



**INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.**

I. A. R. I. 6.

MGIPC-81-51 AR/57-34-58-5,000.

PROCEEDINGS
OF THE
ROYAL SOCIETY OF EDINBURGH

PROCEEDINGS
OF
THE ROYAL SOCIETY
OF EDINBURGH

Section A (Mathematical and Physical Sciences)

VOL. LXIII

1949-1952

PUBLISHED BY
OLIVER & BOYD
EDINBURGH: TWEEDDALE COURT
LONDON: 98 GREAT RUSSELL STREET, W.C. 1

1952

CONTENTS

NO.		PAGE
*1.	The Adventures of an Hypothesis. By James Kendall, M.A., D.Sc., F.R.S., P.R.S.E. (With Two Text-figures.) Issued separately May 17, 1950, .	1
*2.	The Stability of Solutions of Non-linear Difference-differential Equations. By Professor E. M. Wright, University of Aberdeen. Issued separately May 17, 1950,	18
3.	The Simplest Form of Second-Order Linear Differential Equation, with Periodic Coefficient, having Finite Singularities. By Enzo Cambi. <i>Communicated by</i> Professor A. C. Aitken, F.R.S. (With Two Text-figures.) Issued separately June 16, 1950,	27
*4.	Studies in Practical Mathematics. V. On the Iterative Solution of a System of Linear Equations. By A. C. Aitken, D.Sc., F.R.S., Mathematical Institute, 16 Chambers Street, Edinburgh, 1. Issued separately June 16, 1950,	52
5.	Les transformations asymptotiquement presque périodiques discontinues et le lemme ergodique. (Première Note.) Par Maurice Fréchet, Hon.F.R.S.E., Université de Paris, à la Sorbonne. <i>Communicated by</i> Sir Edmund Whittaker, F.R.S. Issued separately June 16, 1950,	61
6.	Unbiased Statistics with Minimum Variance. By A. Bhattacharyya, Statistical Laboratory, Calcutta. <i>Communicated by</i> Professor A. C. Aitken, F.R.S. Issued separately June 16, 1950,	69
7.	Parallel Planes in a Riemannian V_n . By H. S. Ruse, The University, Leeds. Issued separately June 16, 1950,	78
*8.	A Further Note on a Problem in Factor Analysis. By D. N. Lawley, M.A., D.Sc., University of Edinburgh. Issued separately June 16, 1950,	93
*9.	A Measurement of the Velocity of Light. By R. A. Houstoun, M.A., D.Sc., F.Inst.P., Natural Philosophy Department, University of Glasgow. (With Four Text-figures.) Issued separately June 16, 1950,	95
*10.	The Reciprocity Theory of Electrodynamics. By H. S. Green and K. C. Cheng, University of Edinburgh. <i>Communicated by</i> Professor Max Born, F.R.S. Issued separately May 14, 1951,	105
*11.	Application of Relaxation Methods to Compressible Flow past a Double Wedge. By A. R. Mitchell, Ph.D., and D. E. Rutherford, Dr.Math., D.Sc., United College, University of St Andrews. (With Seven Text-figures.) Issued separately May 14, 1951,	139

NO.		PAGE
*12.	Clebsch-Aronhold Symbols and the Theory of Symmetric Functions. By H. W. Turnbull and A. H. Wallace, The University, St Andrews. Issued separately May 14, 1951,	155
*13.	Studies in Practical Mathematics. VI. On the Factorization of Polynomials by Iterative Methods. By A. C. Aitken, D.Sc., F.R.S., Mathematical Institute, University of Edinburgh. Issued separately May 14, 1951,	174
*14.	Experiments in Diffraction Microscopy. By G. L. Rogers, M.A., Ph.D., Department of Physics, University College, Dundee, Angus. <i>Communicated by</i> Professor G. D. Preston. (With Two Plates and Nine Text-figures.) Issued separately February 9, 1952,	193
*15.	Theorems on the Convergence and Asymptotic Validity of Abel's Series. By A. J. Macintyre and Sheila Scott Macintyre, University of Aberdeen. Issued separately February 9, 1952,	222
*16.	Some Continuant Determinants arising in Physics and Chemistry—II. By D. E. Rutherford, D.Sc., Dr.Math., United College, University of St Andrews. Issued separately February 9, 1952,	232
*17.	The Sargent Diagram for the Electron-capture Process, and the Disintegration Energies of Heavy β -emitters. By N. Feather, Department of Natural Philosophy, University of Edinburgh. Issued separately February 9, 1952,	242
*18.	The Elementary Proof of the Prime Number Theorem. By E. M. Wright, University of Aberdeen. Issued separately May 30, 1952,	257
19.	A Generalization of the Classical Random-walk Problem, and a Simple Model of Brownian Motion based Thereon. By G. Klein, Ph.D., Birkbeck College, London, now at Université Libre de Bruxelles. <i>Communicated by</i> Dr R. Fürth. Issued separately May 30, 1952,	268
20.	On the Estimation of Variance and Covariance. By E. H. Lloyd, Imperial College, London. <i>Communicated by</i> Professor H. Levy. Issued separately May 30, 1952,	280
21.	The Statistical Theory of Stiff Chains. By H. E. Daniels, M.A.(Cantab.), Ph.D.(Edin.), Statistical Laboratory, University of Cambridge. <i>Communicated by</i> Professor A. C. Aitken, F.R.S. Issued separately May 30, 1952,	290
*22.	Artificial Holograms and Astigmatism. By G. L. Rogers, M.A., Ph.D., Department of Physics, University College, Dundee. <i>Communicated by</i> Professor G. D. Preston. (With One Plate and Four Text-figures.) Issued separately August 23, 1952,	313
*23.	Studies in Practical Mathematics. VII. On the Theory of Methods of Factorizing Polynomials by Iterated Division. By A. C. Aitken, D.Sc., F.R.S., Mathematical Institute, University of Edinburgh. Issued separately August 23, 1952,	326

NO.		PAGE
*24.	The Solution of a Functional Equation. By A. H. Read, M.A., United College, University of St Andrews. <i>Communicated by</i> Dr D. E. Rutherford. Issued separately August 23, 1952,	336
*25.	The First Chemical Society, the First Chemical Journal, and the Chemical Revolution. By James Kendall, M.A., D.Sc., LL.D., F.R.S., P.R.S.E. (With Two Plates.) Issued separately October 20, 1952,	346
*26.	The Normal Penetration of a Thin Elastic-Plastic Plate by a Right Circular Cone. By J. W. Craggs, Ph.D., University College, Dundee. <i>Communicated by</i> J. M. Jackson, Ph.D. (With Four Text-figures.) Issued separately October 20, 1952,	359
*27.	The Rotational Field behind a Bow Shock Wave in Axially Symmetric Flow using Relaxation Methods. By A. R. Mitchell, Ph.D., and Francis McCall, B.Sc., United College, University of St Andrews. <i>Communicated by</i> Dr D. E. Rutherford. (With Five Text-figures.) Issued separately October 20, 1952,	371
*28.	A Molecular Sum Rule. By D. ter Haar, Department of Natural Philosophy, University of St Andrews. Issued separately December 4, 1952,	381
*29.	The First Chemical Society, the First Chemical Journal, and the Chemical Revolution (Part II). By James Kendall, M.A., D.Sc., LL.D., F.R.S., P.R.S.E. Issued separately December 4, 1952,	385

* The thanks of the Society are due to the Carnegie Trust for the Universities of Scotland for grants towards the costs of the illustrations, tables, etc. in these papers

PROCEEDINGS
OF THE
ROYAL SOCIETY OF EDINBURGH

Section A (Mathematical and Physical Sciences)

VOLUME LXIII

I.—**The Adventures of an Hypothesis.*** By James Kendall,
M.A., D.Sc., F.R.S., P.R.S.E. (With Two Text-figures)

(Address of the President at a Meeting held on December 5, 1949)

(MS. received October 27, 1949)

I. ORIGIN AND RISE

IN the year 1815 an anonymous article appeared in Thomas Thomson's *Annals of Philosophy* (1) entitled "On the Relation between the Specific Gravities of Bodies in their Gaseous State and the Weights of their Atoms". Its introductory paragraph illustrates the hesitancy of the writer in its exposition: "The author of the following essay submits it to the public with the greatest diffidence; for although he has taken the utmost pains to arrive at the truth, yet he has not that confidence in his abilities as an experimentalist as to induce him to dictate to others far superior to himself in chemical acquirements and fame. He trusts, however, that its importance will be seen, and that some one will undertake to examine it, and thus verify or refute its conclusions. If these should be proved erroneous, still new facts may be brought to light, or old ones better established, by the investigation; but if they should be verified, a new and interesting light will be thrown upon the whole science of chemistry."

The general conclusion reached from the experimental work that constitutes the main body of the paper—the results of a great many other investigators are also collected in four tables—is rather indefinitely and tentatively formulated, but it is evident what the author wishes to suggest

* Assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland. This Address will also be published in *Proceedings B*.

he has proved, namely, that the atomic weights of all other elements are exact multiples of the atomic weight of the lightest element, hydrogen. Examining his data critically, one has grave doubts whether this conclusion was really justified. True, the work is of higher accuracy than that of Dalton, whose values for carbon and oxygen differed by more than ten per cent. from the correct figures when he stated in 1810 (2): "An atom of carbonic oxide consists then of one of carbone or charcoal, weighing 5.4, and one of oxygen, weighing 7." (These figures must be doubled to bring them into correspondence with the basis hydrogen = 1.) Dalton established his atomic theory on a very inadequate experimental foundation, and Thomson's anonymous correspondent followed Dalton's example in drawing inspired deductions from insecure premises. He finds, for instance, that the specific gravity of oxygen is *just* 16 times that of hydrogen, and the specific gravity of nitrogen *just* 14 times, but he arrives at these values by treating atmospheric air as a compound constituted by bulk of four volumes of nitrogen and one volume of oxygen! Consider, moreover, his remark: "There is every reason for concluding that the specific gravity of chlorine does not differ much from 2.5 [atmospheric air being 1.000]. On this supposition, the specific gravity of chlorine will be found exactly 36 times that of hydrogen." No element, it is obvious, could fail to fit into a scheme of such flexibility.

Thomson, however, was a firm believer from the beginning. In an account of the improvements in physical science during the year 1815 he wrote (3): "A very important paper was published in a late number of the *Annals of Philosophy* on this subject [the atomic theory]. Though the paper in question is anonymous, several circumstances enable me to fix with considerable certainty on the author; but as he chuses to remain for the present concealed, I do not consider myself as at liberty to mention his name." Thomson had already been engaged for several years in atomic weight determinations, and he now embarked enthusiastically on a comprehensive series of experimental researches designed to establish the validity of the idea advanced by his contributor.

This same contributor broke into print again also, early in 1816, with a second anonymous article in the *Annals of Philosophy* (4). It was only three pages in length, and it purported to be merely the correction of a mistake in the previous paper ("an oversight which influences some of the numbers in the third table"), but its sting lay in its tail: "If the views we have ventured to advance be correct, we may almost consider the *πρώτη ἰλη* of the ancients to be realised in hydrogen." In other words, not only are the atomic weights of heavier elements exact multiples of that of hydrogen, but the atoms of heavier elements are to be regarded as aggregates of

hydrogen atoms. Hydrogen is thus promoted to the rôle of the primordial substance from which, according to the Greek philosophers, all other material substances in the universe are derived. A few months later Thomson (5) disclosed the fact that this bold speculation emanated from the brain of Dr William Prout. It has established itself in the history of science as *Prout's Hypothesis*.

Who was Dr William Prout? He was born on 15th January 1785 at Horton, in Gloucestershire, where his family had been landed proprietors for several generations. His early education was neglected but, like many other Englishmen before and since, he remedied the deficiency by betaking himself to the University of Edinburgh, where he obtained the degree of M.D. in 1811. After graduation he carried on a medical practice in London, devoting his spare time to research. He became a pioneer investigator in the field of physiological chemistry, and was elected a Fellow of the Royal Society of London in 1819. He contributed thirty-four papers in all, mostly upon medical subjects, to various journals. In his later years he became, owing to deafness, somewhat of a recluse. He died on 9th April 1850.

One of Prout's discoveries in medicine—the presence, in the gastric fluid, of free hydrochloric acid, a most important factor in digestion—was of primary significance, but his real claim to fame rests upon the remarkable hypothesis of the unity of matter which he proposed with such a combination of diffidence and audacity at the early age of thirty. Though he took no active part in its further development, the pious hope which he expressed that others, by testing its truth, would throw a new and interesting light on the whole science of chemistry has been amply fulfilled.

In Great Britain the majority of chemists, following the lead of Thomas Thomson of Glasgow, accepted the hypothesis of Prout as absolute truth. Dalton, however, held aloof: "No man can split an atom" was his dogma, and he could not be induced to abandon it. On the Continent, opinion was generally adverse. Berzelius, the last dictator of chemistry, had decided from his researches that no simple relation existed between the weights of atoms of different elements, and as long as Berzelius lived that view almost universally prevailed. Everywhere, however, the urge to establish more accurate values for atomic weights, and so put the hypothesis to strict trial, led to rapid improvements in analytical technique, and data of greater reliability steadily accumulated.

2. DECLINE AND FALL

For a long time the bulk of the evidence continued favourable. Turner (6) in 1833 and Penny (7) in 1839 did obtain results which led them to

conclude that the hypothesis was not exact, but the disagreement of certain of their own values weakened their objections. Dumas and Stas (8), from a very careful synthesis of carbon dioxide, deduced in 1841 a figure for the atomic weight of carbon that agreed perfectly. New syntheses of water by Dumas (9) in 1843, confirmed by the work of Erdmann and Marchand, were also in complete accordance.

Simultaneously, however, Marignac (10, 11) published the results of a whole series of researches undertaken for the purpose of submitting Prout's Law (it was now beginning to be granted this superior status) to a new and rigid examination. Marignac declared himself to be of the firm opinion that, taking into account the extreme difficulty of attaining experimental values of absolute exactness, we cannot consider as contrary to Prout's Law his own figures, determined to three places of decimals, for silver, potassium, bromine, iodine and nitrogen. Chlorine, however, presented a clear exception; its atomic weight, 35.456, might conceivably be rounded off to 35.5, but an integral value was absolutely excluded. To rescue Prout's Law in this emergency, Marignac suggested the adoption of a basic unit of 0.5, namely half the weight of the hydrogen atom.

It is of interest to note that Prout himself had anticipated this suggestion. In a letter to Daubeny (12) in 1831, he mentioned that he saw no reason why bodies still lower in the scale than hydrogen might not exist, of which other bodies might be multiples, without being actually multiples of the intermediate hydrogen. The substitution of a totally unknown element for hydrogen as primordial matter was a device, however, which devalued Prout's Hypothesis very drastically. When in 1859 Dumas (13), to account for further discrepancies that had become apparent, proposed a reduction of the standard to 0.25, it was felt that such a procedure might be continued indefinitely, depriving the hypothesis of any permanent significance.

The supreme crisis, resulting in what looked like complete collapse, came in 1860, when Dumas' former collaborator, Jean Servais Stas (14), submitted an impressive memoir to the Royal Academy of Belgium, summarising the results of twenty years' researches on atomic weights. What regret he experienced in communicating these results may be gauged from the following extract from a conversation that he had many years later with a young compatriot, Leo Hendrick Baekeland (15): "In my youth I was an ardent believer in the unity of matter as expounded by Prout. I was so well convinced about his theory that I became eager to furnish additional proofs by redetermining more accurately the atomic weights of those elements where the atomic weight numbers were not an even multiple of hydrogen. I simply imagined that more careful determinations would have eliminated these irregularities. But the more I

worked, the more I perfected my methods, the more I eliminated any errors of experimentation, so much the more did my results contradict my dearest hopes. Finally I had to admit that I was beaten and had spent the most important part of my life in killing my first love as a theory."

It must be granted that Stas wielded the axe very sympathetically. He accorded eloquent homage to the rare penetration of Prout, and emphasised the immense importance of his idea of the unity of matter from the philosophical point of view, but he proceeded to administer the *coup de grâce* as follows: "All doubt has vanished from my mind. I have reached the complete conviction, the entire certainty, as far as certainty can be attained on such a subject, that Prout's Law, with all the modifications due to M. Dumas, is nothing but an illusion, a pure hypothesis expressly contradicted by experiment. Chemists after examining the work which I have the honour of now presenting in detailed analysis to the Academy, if they can cast away their prejudices and mental prepossessions, and place their reliance on experiment, will soon share my conviction, namely, that there does not exist a common divisor for the weights of the elements which unite to form all definite compounds."

Chemists did, indeed, share the conviction of Stas, and Prout's Hypothesis of the unity of matter was decently buried for the next forty years. There was just one final twitching of the corpse before interment. Marignac (16), in a commentary upon the memoir of Stas, after paying tribute to his splendid experimental work, remarked: "It may perhaps be found astonishing that I do not entirely agree with the conclusions of M. Stas that we must consider Prout's Law as a pure illusion and regard the undecomposable bodies of our globe as distinct entities. . . . While preserving the fundamental principle of that law, that is to say, while adopting the hypothesis of the unity of matter, could we not make for example the following supposition, to which I do not otherwise attach importance except as showing that we might explain the discordance that appears to exist between the results of observation and the immediate consequences of that principle? Could we not suppose that the cause (unknown but probably different from the physical and chemical agencies familiar to us) which has determined certain groupings of the atoms of the sole primordial matter so as to give rise to our chemical atoms, by impressing on each of these groups a special character and particular properties, should not at the same time have exercised an influence on the manner according to which these groups of primordial atoms would obey the law of universal attraction in such wise that the weight of each group might not be exactly the sum of the weights of the primordial atoms composing it?"

In my own opinion, let me interpolate, this supposition of Marignac, to which he himself attached no intrinsic importance and which was entirely disregarded by his contemporaries since it contravened the fundamental law of the conservation of mass, constitutes the most astounding episode in the whole narrative. Marignac actually, as will appear later, anticipated the discovery of the packing effect in the atomic nucleus! He made, however, merely a stab in the dark which it was impossible to follow up immediately; fifty years had to elapse before definite proof could be adduced that his supposition was valid.

One indirect line of evidence, which was available during the interval, provoked sporadic speculation, but its real significance was never noted. Purely for convenience, chemists had gradually changed over from the basis hydrogen = 1 to the basis oxygen = 16 for the expression of atomic weights. Hydrogen combines directly with very few other elements, whereas oxygen combines directly with almost all, and the accurate determination of the hydrogen-oxygen ratio in water is a matter of extreme difficulty. It was therefore necessary, under the hydrogen standard, to change the atomic weight of nearly every element on the list each time some ingenious person redetermined the hydrogen-oxygen ratio a little more exactly. So long as hydrogen retained its rôle of primordial matter, this awkwardness might be endured, but once Prout's Hypothesis was discarded, chemists naturally turned to the more rigid oxygen basis.

The difference is only slight—on the scale oxygen = 16, hydrogen becomes 1.0078—but it entails a very curious consequence. An extraordinary proportion of the atomic weights, with oxygen = 16, fall so close to being integral that the nearest round numbers are exact enough for ordinary use. This is not the case when hydrogen = 1 is made the basis, as the following typical section of the periodic system, which includes all the elements from fluorine to chlorine, will demonstrate.

Element	Atomic Weight (O = 16)	Atomic Weight (H = 1)
Fluorine	19.00	18.86
Neon	20.183	20.030
Sodium	22.997	22.822
Magnesium	24.32	24.13
Aluminium	26.97	26.76
Silicon	28.06	27.84
Phosphorus	30.98	30.75
Sulphur	32.06	31.82
Chlorine	35.457	35.195

Here is indeed an anomaly. By the abandonment of Prout's Hypothesis the chemists came right back to another form of it, since it has been

calculated that the odds are billions to one against such an agglomeration of atomic weights in the immediate vicinity of integral values occurring merely by chance. There are still some flagrant exceptions, of course, such as magnesium and chlorine, but it is obvious that the atomic weight of hydrogen is essentially out of step with those of other elements. The logical deduction, however, that the mass of the hydrogen atom is altered by an approximately constant fraction when it is incorporated into heavier atoms, was never drawn. Marignac's supposition, which afforded the key to the difficulty, had passed into oblivion.

3. RESURGENCE

The turn of the century brought much more cogent reasons for regarding chemical elements as possessing a common basis. Once it was established that uranium, by a sequence of radioactive disintegrations in which positively charged helium atoms and electrons were ejected, transmuted into other elements such as radium and lead, the idea of the unity of matter came again to the fore. The approaching centenary of Prout's Hypothesis was fittingly celebrated by Rutherford's formulation in 1911 of the theory of the nuclear atom.

According to this theory, the hydrogen atom consists of a minute positively charged nucleus or proton, responsible for almost its entire mass, round which rotates a negatively charged unit or electron, responsible for almost its entire volume. The atoms of all other elements comprise a steadily increasing number of exterior electrons, with nuclei that contain, besides an equal number of free protons, a number of proton-electron pairs, "collapsed hydrogen atoms" or neutrons. All heavier atoms are therefore, essentially, aggregates of hydrogen atoms, and Prout's Hypothesis is rehabilitated.

Two points demand discussion: how can the abnormality in the atomic weight of hydrogen be explained, and how can other recalcitrant elements such as chlorine be brought into line? The answer to the first question has already been hinted at above, and will be developed in detail later. The solution of the second problem may be very briefly indicated.

Chlorine, with an atomic weight of 35.457, has been proved to be a mixture of two types of atoms, with masses 35 and 37 respectively. Their structures are: ^{35}Cl = 17 exterior electrons; 17 free protons and 18 neutrons in the nucleus; ^{37}Cl = 17 exterior electrons; 17 free protons and 20 neutrons in the nucleus. All naturally occurring chlorine contains the two types in the ratio required to make the apparently abnormal atomic

weight the average of two normal aggregates. Many other elements are similarly *isotopic*—even the two standards, hydrogen and oxygen, contain minute traces of heavier atoms—but *all* the isotopic varieties of *every* element fit perfectly into Rutherford's scheme.

It was thought at first that all atomic masses might be fundamentally integral on the oxygen=16 basis, with hydrogen (1.0078) as the sole exception. More recent determinations, which have now reached the phenomenal accuracy of five places of decimals, have shown that this is not the case—there is a packing factor in the nucleus inducing small, regularly varying deviations.

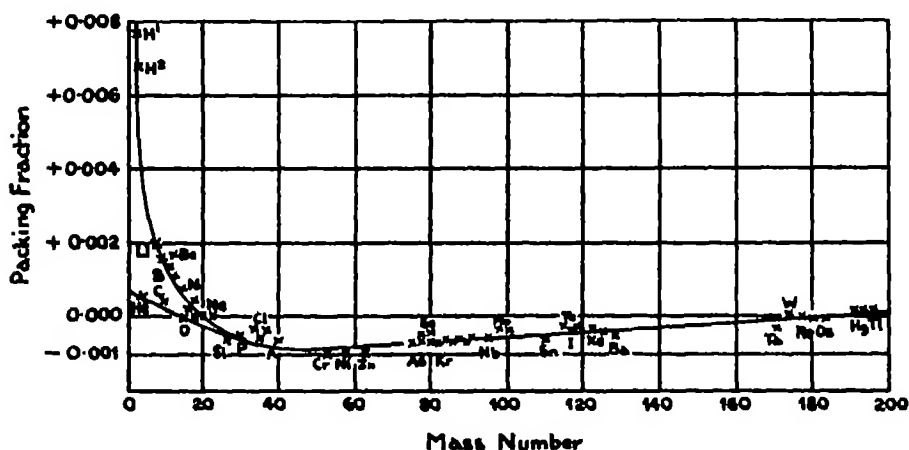


FIG. 1.—Variation of packing effect with atomic mass.

Inside the very minute nuclei of all atoms except the simplest, hydrogen, protons and neutrons are packed so closely together that their electromagnetic fields interfere and a fraction of the combined mass is destroyed. The helium atom, for example, contains the same total material as four hydrogen atoms, but its mass is *not* $4 \times 1.0078 = 4.0312$; it is only 4.003. The proportionate influence of this packing effect on atomic mass for still more complex nuclei is almost constant, but not exactly so. This may best be illustrated by comparing all atomic masses with that of the mass-spectrograph standard, ^{16}O . Fig. 1 shows the curve obtained when mass number (the *total* number of protons in the nucleus) is plotted against what Aston has called the packing fraction (the main gain or loss in mass per proton when the nuclear packing is changed from that of ^{16}O to that of each other atomic species).

Several points of interest emerge from this curve. The initial, steeply descending sections (the curve splits into two distinct branches for the

earlier elements, following those of odd and even atomic number respectively) indicate that the masses of atoms lighter than ^{16}O are *slightly higher*—in the single case of hydrogen significantly higher—than whole numbers. Through a long range of elements in the central part of the diagram the packing fraction is negative, so that atomic masses in this interval are *slightly lower* than whole numbers. In the region of the rare earths the curve crosses the zero line once more, and atomic masses are again above integral values. A more accurate graph, on a larger scale, for these latter sections is given in fig. 2.

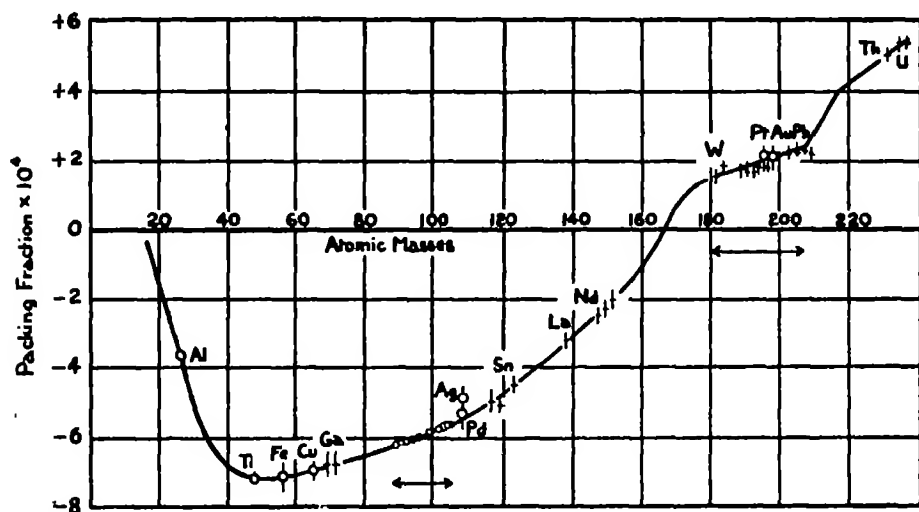


FIG. 2.—Dempster's packing fraction curve (1938).

Elucidation of these variations is obtained by the application of the theory of relativity. What we have lost in mass through nuclear packing we have gained in energy. Mass and energy are not distinct phenomena, but interconvertible; matter is potential energy, energy is potential matter. The law of conservation of mass holds within our limits of weighing, for all ordinary chemical reactions, only because the energy changes involved are extremely small. Energy changes in reactions "within the atom", however, first demonstrated in radioactive disintegrations, may be enormous. Correlating our old laws of conservation of mass and conservation of energy into one wider law—*conservation of mass-energy*—we can calculate what change in mass corresponds to any particular energy change, and *vice versa*. We find that the mass change in ordinary chemical reactions is infinitesimal, and that even in radioactive disintegrations it is still minute. Conversely, if the mass change is appreciable,

as in the case of hydrogen-helium, the energy change involved in the rearrangement of the protons and electrons must be stupendous.

It is very significant that the minimum in the packing fraction curve lies in the neighbourhood of iron, one of the most abundant elements in nature. Atoms in the region of this minimum should be most stable, whereas atoms at the two extremes of the diagram should be susceptible to highly exothermic transformations to other types. Such transformations, in the case of normal radioactive disintegrations, carry us only short distances along the curve at the right of fig. 2. A much longer jump, however, has recently been achieved by the fission of uranium, which renders the utilisation of atomic energy possible in peace as well as in war. Atomic fusion of the lighter elements would obviously release even greater amounts of energy per unit weight of matter transmuted. This has not yet been effected on a major scale on our planet (although, by bombarding a target of lithium with protons of 700,000 electron volts energy, Cockroft and Walton (17) induced the reaction, ${}^7\text{Li} + {}^1\text{H} \rightarrow {}^4\text{He} + {}^4\text{He}$, where the energy of the ejected α -particles exceeds 8,000,000 electron volts, as early as 1932), but a series of reactions in which the ${}^{12}\text{C}$ atom functions as a catalyst for the transformation of hydrogen into helium (18) is regarded by astrophysicists as offering the first adequate explanation for the source of the heat of the sun throughout the lifetime of our solar system.

4. A VISION OF THE FUTURE

It may seem that I have strayed from Prout's Hypothesis in the preceding paragraphs, but I have expatiated on the topic of energy because I wish to conclude this address by presenting for your consideration a comprehensive extension of Prout's Hypothesis in which the fundamental fact that matter and energy are now practically synonymous is essentially involved. I shall restrict my exposition to the barest outline, since I am not qualified to delineate it in detail, and I leave it to my colleagues in biology, medicine and philosophy to elaborate the evidence on sections which I may cover inadequately or omit entirely, and to correct any errors that I may inadvertently make.

Let me first recapitulate the present position with respect to matter. Prout's first postulate, that the atomic weights of all other elements are exact multiples of that of hydrogen, has had to be discarded, but a perfectly logical explanation for all the divergences therefrom is now available. His second and more significant postulate, the unity of matter, has been triumphantly resuscitated in Rutherford's picture of the nuclear atom in terms of protons and electrons. (Complexities introduced by the discovery

of transitory and intermediate units, such as mesons, positrons, neutrinos, need not be entered into here.)

Next let me return for a moment to energy. The interconvertibility of all forms of matter has only recently been indubitably proved, but the interconvertibility of all forms of energy was established more than a hundred years ago. In 1900 Planck advanced the hypothesis that, in the same way as matter was made up of discrete particles or atoms, so energy also existed only in discrete indivisible units or quanta. The application of Planck's *quantum theory* has been just as fruitful in its results as was the application of Dalton's atomic theory. In consequence of the "atomicity" of energy, the electron orbits in different atoms are restricted to definite energy-levels, which in the simpler cases can be calculated to correspond *exactly* with experiment.

The mass of the atom varies according to the number of protons and electrons in its composition; the magnitude of the quantum ϵ is expressed by the simple equation $\epsilon = h\nu$, where h is a universal constant (Planck's constant), and ν is the frequency (the reciprocal of the wave-length) of the energy under consideration. Quanta therefore decrease enormously in magnitude as we proceed from energy of short wave-length (X-rays) through the visible spectrum to energy of long wave-length (radio communication). The variation differs from that of atoms only in its continuity.

Apart from this detail, matter and energy stand now on precisely the same footing: the unity of matter is duplicated by the unity of energy. Matter and energy, as described in the preceding section, are also mutually interconvertible, a fact expressed in what we may truly regard as the modern form of Prout's Hypothesis, the law of the conservation of mass-energy with the implications attendant thereupon (see p. 9).

Now, besides matter and energy, there is a third principle which has been recognised as existent in the universe, a principle just as varied in its manifestation as matter and energy themselves, a principle which the philosophers called *mind*, but which we may amplify here to the still more general concept of *life*.

So long as matter and energy were regarded as distinct phenomena there was no reason at all to surmise that life had any fundamental correlation with either, but now that matter and energy have been completely integrated, it is surely opportune to analyse an hypothesis that vital phenomena are also amenable to inclusion in the same category. Such an analysis, indeed, I propose to make, with Prout's Law (it well merits that title again to-day) as my guide. Perhaps I can best make clear the idea that I wish to convey by considering what conclusions scientists of

different periods in the past might logically have reached if they had attempted a similar analysis, with the evidence then at their disposal, before I develop what may be inferred from present knowledge.

Previous to Prout, the line of reasoning would have been simple. Many elements were known, all quite distinct from one another, quite unsusceptible to change. By analogy, the many forms of life might be assumed to be absolutely unrelated, immutable from the time of the creation. Lavoisier's statement of the indestructibility of matter would be paralleled by the statement of the immortality of the soul. Everything, in short, would be in complete accordance with orthodox views. If energy were also brought into consideration, Rumford's experiment on the conversion of mechanical to thermal energy might have been deemed disturbing; but too little was known at that time about energy to be of any real value.

After the advent of Prout's Hypothesis, however, the argument would have run very differently. All the many varieties of matter are fundamentally related; consequently the assumption would follow, by analogy, that all the many varieties of life are similarly related. Hydrogen, the simplest and primordial form of matter, from which all other forms of matter are derived, would be paralleled by the single cell, the simplest and primordial form of life, from which all other forms of life are derived. This basis for a unitary conception would have been strengthened, around the middle of the nineteenth century, by the recognition of the interconvertibility of all forms of energy. Had biology been a little further advanced, in fact, the line of reasoning would have accorded exactly with that of the early evolutionists, culminating in Herbert Spencer's philosophical doctrine of general development in the universe from homogeneous and simple to heterogeneous and complex.

After 1860, any aspiring co-ordinators would have found themselves in a quandary. With the downfall of Prout's Hypothesis, the various forms of matter were once more regarded as essentially distinct and incommutable. Interconvertibility of the various forms of energy, however, remained uncontested. It is obvious that no conclusions, by analogy, regarding the various forms of life could possibly be reached.

Our present basis for inquiry, to which I finally revert, is vastly more promising. Matter and energy have been not merely brought into line but amalgamated, providing an harmonious system in which it is plausible to postulate that life may also be incorporated. Carrying our analogy to the extreme would involve the hypothesis of the existence of a comprehensive future law which might be termed the law of the conservation of mass-energy-life; but the enunciation of such an hypothesis at present

would be altogether premature, since its implications would be far too indefinite and there would be no means of testing its validity. The reasons for our inadequacy may be explained very briefly.

Although we know that all matter is, by ultimate analysis, reducible to energy, it is only in the case of a few of the simplest forms of matter that we are able to express their properties quantitatively on an energy basis (see p. 11). With increasing complexity of the atom, most relationships—X-ray spectra provide a welcome exception—soon become too intricate for calculation. Much more is this so when we proceed to compounds, particularly to large molecules. While we recognise that constitutionally these are still ultimately reducible to energy, we recognise also the utter impossibility of establishing their properties purely in terms of energy until enormous advances have been made in our present knowledge. We continue, therefore, to investigate the "properties of matter" and to consider them as such in the meantime.

If now we postulate, by analogy, that all vital phenomena may be ultimately expressed in terms of matter and energy, we immediately find ourselves in a similar predicament. These phenomena relate to units so complex that the largest synthetic organic molecules are simple in comparison. Consequently, although vital processes may be successfully described in physico-chemical terms up to a certain point, we soon reach limits where explanation on such a basis becomes incomplete or quite impracticable. Pending further progress in our knowledge which will push forward these limits, therefore, we continue to investigate the "properties of living material" and to consider the processes involved as "life processes".

The gap here is comparable, indeed, with that which existed between matter and energy in 1860. The physicists of the present century gradually narrowed that gap until it has been securely bridged, but even if a second Marignac were to hazard an inspired guess as to how the link of matter and energy with life might be fully established, we should be as little qualified to interpret or verify it as the contemporaries of Marignac were to interpret or verify his prognostication of the packing effect (p. 6).

Within restricted bounds, nevertheless, I believe that the analogy that I have postulated does enable us to formulate deductions of immediate interest. A brief, and necessarily incomplete, presentation of one such topic is attempted in the final section of this address.

5. A PRESENT GLIMPSE OF THE VISION OF THE FUTURE

In a myriad different species of living bodies we can definitely distinguish a material factor, the number and nature of their respective cell

chromosomes, reminiscent on a larger scale of the arrangements of atoms and groups of atoms in a myriad different complex chemical compounds. Associated with this material factor so long as life persists, an energy factor is demanded, which used to be designated vital energy. It is now generally recognised (19) that this is not inherently distinct from other forms of energy in any of its many manifestations susceptible to experimental investigation, ranging from mechanical and thermal energy to the electrical discharges along nerve-fibres (thought impulses) established by recent studies in electro-encephalography. Whether this association of matter and energy is sufficient to account for more abstract features of vital phenomena (*e.g.*, instinct, intelligence, reasoning, etc.) is a highly controversial question which lies completely beyond the bounds of present inquiry.

For the continuance of life in any specimen thereof, the maintenance of stable conditions among the material and energy factors is essential; these conditions become more complex the greater the complexity of the organism. Some of the limitations (*e.g.*, temperature range, chemical environment) have been definitely determined; with respect to others, however, our knowledge remains very elementary. A typical example of the present stage of experimental progress on one vital problem—the propagation of living tissue *in vitro*—may be cited. While it has been found possible to maintain simple tissues, such as chicken embryo tissue, growing continuously after removal from the body for decades in carefully adjusted synthetic surroundings, the conditions favouring and disfavouring the persistence of more complex aggregates provide innumerable problems still in the initial stages of attack.

One most important fact, nevertheless, has been experimentally proved. Many independent investigations have shown that mutations in living species can be artificially induced both by drastic change in energy environment (exposure to short-wave irradiation) and by drastic change in material environment (exposure to chemical action). Mention may be made, in particular, of the experiments of Muller (20) with X-rays and of Auerbach and Robson (21) with mustard gas. These investigations demonstrate beyond doubt that by means of matter and by means of energy we can change the character of the chromosomes, or of the genes that they carry, and so vary the form of life. The changes in the chromosomes are often, though probably not always, visible, and are structural in nature, consisting of inversions, translocations and deletions. Bacteria, unicellular organisms with an internal structure so delicate that they do not possess the obvious chromosomes of higher forms of life, also exhibit changed properties on exposure to drugs or to ultra-violet light.

This convertibility of one form of life to another by means of matter

and energy, limited as it is at the moment, offers wide possibilities for future developments. Reasoning by pure analogy from similar developments in the fields of matter and energy in the past, I submit now what I believe to be a legitimate hypothesis on the basis of current knowledge—the co-ordination of constitutional changes in mass-energy-life.

Let us consider the cell as our unit, analogous to a complex molecule, and consider the chromosomes as corresponding to constituent groups thereof. On this basis, if parallelism does prevail, the various types of mutagenic reactions should duplicate the various types of chemical reactions. The following classification suggests that this is substantially true:—

(a) The majority of spontaneous mutations, resulting in the gradual evolution of new species, may be regarded as equivalent to ordinary chemical changes with very low reaction rates. (Any rapid reaction, it may be noted, would have been long ago completed.) Like ordinary chemical reactions, these spontaneous mutations should (and do) exhibit large temperature coefficients.

(b) A small fraction of spontaneous mutations, induced by the action of stray radiations, may be regarded as equivalent to photo-chemical reactions. As such, they should not normally exhibit any temperature coefficient. This question cannot be tested directly, but its presumable validity is indicated in (d) below.

(c) Artificial mutations induced by chemical agents such as mustard gas may be regarded as equivalent to catalytic reactions. The mustard gas or other agent functions as a catalyst, increasing the natural rate of reaction very significantly. Spontaneous mutations may possibly be due, in whole or part, to the natural occurrence of minute amounts of such catalysts in living material. Whether this is so or not, chemically induced mutations should show a large temperature coefficient. This important point has not yet been investigated experimentally.

(d) Artificial mutations induced by irradiation may be regarded as equivalent to photo-catalytic reactions. Their rate is proportional to the intensity of the radiation, which is here enormously increased over that in (b) above. They should not, as photo-catalytic reactions, exhibit any temperature coefficient (although subsequent chemical reactions may introduce a small temperature factor), and experiment shows that this is in fact the case.

Most of the changes resulting from reactions of the above types, inasmuch as they affect only a small part of a complicated "molecule", may be expected to be limited in their influence on the properties of the organisms as a whole. New forms of life manufactured thereby, as in the

case of new strains of *Penicillium chrysogenum* obtained by X-ray irradiation, will therefore in general exhibit only minor differences from their natural source. Major changes may be anticipated to be rare, and in most cases to be harmful, recessive, or lethal.

Nevertheless, occasional modifications of extensive character must be produced, which are capable of survival and persistence. These more significant redistributions within the cell, necessary for the emergence of essentially new species, presumably all have had a purely chemical basis. In such evolutionary developments, both "chromosome fissions" and "chromosome fusions" (compare p. 10) of an abnormal type probably play an important part.

What I have put forward for your consideration is at present, of course, merely an hypothesis, and by no means an entirely novel hypothesis. I trust that you will not think me presumptuous if I dare to cite a precedent. Prout himself, when he first ventured to say in 1816, "We may almost consider the πρώτη ὕλη of the ancients to be realised in hydrogen", followed this up immediately with the words: "an opinion, by the by, not altogether new." This warrants the assumption that the idea was "in the air", and vague foreshadowings may be recognised as existing, in more or less hazy form, in certain passages in the writings of Dalton, Davy and Thomas Thomson (22). In the same way, views on the atomicity of the soul were expressed by the Greek philosophers, and the equation of vital processes with matter and energy has been so frequently mooted, more or less empirically, in recent years, that many scientists accept it as a truism even beyond present experimental limits. As far as I know, however—I speak open to correction—the numerous arguments advanced and inferences drawn have never been reinforced by the compelling analogy of the historical evolution of the Law of Prout.

I have been forced, like Prout, to be indefinite and tentative, and I should be glad if you would permit me to turn back to the first page of this paper and repeat the first paragraph of his first communication in full. It is quite probable that, in far less than fifty years' time, some second Jean Servais Stas will, on the basis of much wider and more exact evidence, pronounce his conviction that even the co-ordination of mutations in life with mutations in matter and energy is an illusion, a pure hypothesis expressly contradicted by experiment. But I trust I may be pardoned for not excluding the possibility that, still later, some second Rutherford will arise to put into concrete form a universal law (hinted at on p. 12) which will transcend my partial exposition more brilliantly than our present Law of Prout transcends the hypothesis which he so hesitantly advanced in 1816.

Much of the material in the earlier part of this address is taken from the historical introduction to Alembic Club Reprint, No. 20 (E. & S. Livingstone, Edinburgh), which reproduces under the title *Prout's Hypothesis* the original papers of Prout, Stas and Marignac. Since several reviewers have regretted the fact that this historical introduction was unsigned, it may be proper for me to disclose now that it was the joint effort of Dr Leonard Dobbin (the senior Fellow of this Society, still scientifically active in the sixty-ninth year of his membership and the ninety-second year of his age) and myself. My sincere thanks are due to Dr Dobbin for his contribution to this odyssey of the Hypothesis of Prout.

REFERENCES TO LITERATURE

- (1) [PROUT], 1815. *Annals of Philosophy*, VI, 321-330.
- (2) DALTON, 1810. *New System of Chemical Philosophy*, Part II, 377.
- (3) THOMSON, 1816. *Annals of Philosophy*, VII, 17.
- (4) [PROUT], 1816. *Annals of Philosophy*, VII, 111-113.
- (5) THOMSON, 1816. *Annals of Philosophy*, VII, 343.
- (6) TURNER, 1833. *Phil. Trans.*, CXXIII, 523-544.
- (7) PENNY, 1839. *Phil. Trans.*, CXXIX, 13-33.
- (8) DUMAS and STAS, 1841. *Ann. chim. phys.* [3], 1, 5-38.
- (9) DUMAS, 1843. *Ann. chim. phys.* [3], VIII, 189-207.
- (10) MARIGNAC, 1842. *C.R.*, XIV, 570-573.
- (11) MARIGNAC, 1843. *Bibl. Univ.*, XLVI, 350-377.
- (12) DAUBENY, 1831. *Introduction to the Atomic Theory*, 129-130.
- (13) DUMAS, 1859. *Ann. chim. phys.* [3], LV, 129-210.
- (14) STAS, 1860. *Bull. Acad. Roy. Belg.* [2], X, 208-336.
- (15) KENDALL, 1949. *Chemistry and Industry* (Baekeland Memorial Lecture), XXVII, 67-70.
- (16) MARIGNAC, 1860. *Bibl. Univ.*, IX, 97-107.
- (17) COCKROFT and WALTON, 1932. *Proc. Roy. Soc. London*, A, LXXXVI, 619.
- (18) BETHE, 1939. *Phys. Rev.*, LV, 434.
- (19) HALDANE, 1949. *Philosophy for the Future*, The Macmillan Company, 204.
- (20) MULLER, 1927. *Science*, LXVI, 84.
- (21) AUERBACH and ROBSON, 1947. *Proc. Roy. Soc. Edin.*, B, LXII, 271, 284, 307.
- (22) [DOBBIN], 1932. Alembic Club Reprint, No. 20 (E. & S. Livingstone, Edinburgh), 5.

(Issued separately May 17, 1950)

II.—The Stability of Solutions of Non-linear Difference-differential Equations.* By Professor E. M. Wright, University of Aberdeen.

(MS. received November 23, 1948. Read February 7, 1949)

SYNOPSIS

Poincaré, Liapounoff, Perron and others have proved theorems about the order of smallness, as the independent variable tends to $+\infty$, of solutions of differential equations with non-linear perturbation terms. A similar theory exists for difference equations. By a simple use of transforms, we here extend the theorems, with suitable modifications, to difference-differential equations. The results are an essential step in the development of a general theory of non-linear equations of this type.

1. A DIFFERENCE-DIFFERENTIAL equation is one involving some or all of the functions

$$y^{(\nu)}(x+b_\mu) = \frac{d^\nu}{dx^\nu} y(x+b_\mu) \quad (0 < \mu < m, \quad 0 < \nu < n)$$

of a real variable x , where m, n are positive whole numbers and

$$0 = b_0 < b_1 < \dots < b_m.$$

As "boundary conditions" we suppose assigned the value of $y^{(\nu)}(0)$ for $0 < \nu < n$ and of $y^{(n)}(x)$ in the initial interval $0 < x < b_m$, $y^{(n)}(x)$ being of integrable square in this interval. We regard our equation as an integral equation in the unknown function $y^{(n)}(x)$ and define $y^{(\nu)}(x)$ for $\nu < n$ by

$$y^{(\nu)}(x) = y^{(\nu)}(0) + \int_0^x y^{(\nu+1)}(\xi) d\xi. \quad (1.1)$$

We say that a solution exists for $0 \leq x \leq X$ if there is an integrable $y^{(n)}(x)$ which satisfies the equation for these values of x . In what follows any statement involving μ or ν holds, unless the contrary is stated, for all integral μ, ν such that $0 < \mu < m, 0 < \nu < n$. We use \sum_μ and \sum_ν to denote summation over these ranges.

Several authors have discussed the linear equation with constant coefficients, viz.

$$\Lambda(y) \equiv \sum_\mu \sum_\nu a_{\mu\nu} y^{(\nu)}(x+b_\mu) = v(x), \quad (1.2)$$

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

where $v(x)$ is a known function. References are given in Wright (1949 a) to their work and to applications of special cases of (1.2) to economics, radiology and the theory of control mechanisms. A particularly important case of (1.2) is the homogeneous equation

$$\Delta(y) = 0, \quad (1.3)$$

in which $v(x) = 0$. It is readily verified that $y(x) = e^{sr^x}$ is a solution of (1.3), provided s_r is a root of the transcendental equation

$$\tau(s) \equiv \sum_{\mu} \sum_{\nu} a_{\mu\nu} s^{\nu} e^{b_{\mu}s} = 0. \quad (1.4)$$

In fact, under fairly wide conditions, the general solution of (1.3) is

$$y(x) = \sum_{i_r} P_r(x) e^{s_r x},$$

the summation extending over all the zeroes of $\tau(s)$. Here $P_r(x)$ is a polynomial of degree one less than the order of the zero s_r of $\tau(s)$, and the coefficients in $P_r(x)$ can be expressed fairly simply in terms of the "boundary conditions".

If we replace the fixed numbers $a_{\mu\nu}$ in (1.2) by functions $A_{\mu\nu}(x)$ tending to $a_{\mu\nu}$ as $x \rightarrow +\infty$, we have a "perturbed" equation. The relationship between the solutions of this new equation and those of (1.2) was found in Wright (1948), which contains references to work by Poincaré, Perron and Bochner on the corresponding problem for pure differential and pure difference equations.

Here I consider another form of perturbed equation, viz. that obtained by adding to the left-hand side of (1.3) a non-linear function of the $y^{(n)}(x + b_{\mu})$ for $\nu < n$, this function being of smaller order than the linear terms when these last are themselves small. Under suitable conditions I prove that, if the $a_{\mu\nu}$ are such that all solutions of (1.3) tend to zero as $x \rightarrow +\infty$, and if the values of $y^{(n)}$ given in the boundary conditions are sufficiently small, then the $y^{(n)}$ in the non-linear equation remain small and, in fact, tend to zero as $x \rightarrow +\infty$. Poincaré (1892), Liapounoff (1907), Perron (1928, 1929) and many others have found similar results for pure differential and pure difference equations. Bellman (1947) is the latest to write on the topic and gives further references.

An obvious solution of the equation which I consider here is the zero solution $y(x) = 0$. My present result is equivalent to the statement that, under suitable conditions, the zero solution is stable; that is, a solution initially small enough approaches the zero solution as $x \rightarrow +\infty$.

In Wright (1946) I stated that part of the theory of the non-linear equation could be developed in three stages. In the first we seek initial

conditions which ensure that a solution tends to zero as $x \rightarrow +\infty$. In the second we show under suitable hypotheses that every solution tending to zero is $O(e^{-cx})$ for some $c > 0$. In the third, with which I was then concerned, we find an asymptotic expansion for a solution which is $O(e^{-cx})$. In the present paper, then, I solve the first stage (as it happens, the first and second stages together) for a fairly wide class of equations.

There are at least two methods available to investigate problems of this type. One of these is based on certain results of Pitt (1944, 1947). Elsewhere (Wright, 1949 *b*) I apply it to consider the stability of solution of a more general class of equations in which $\Lambda(y)$ is replaced by

$$\sum_k \int_0^b y^{(n)}(x - \xi) dk, (\xi)$$

and the non-linear perturbation by an appropriate functional of y . By this means we can also find results corresponding to those of Cesari (1939), Levinson (1946) and Weyl (1946) on the boundedness of solutions of differential equations. But it seems worth while to present here a short proof by another method (a little simpler if a little less powerful) for the case of the difference-differential equation. For the latter equation, the existence and uniqueness of a solution are trivial corollaries of well-known results in the theory of differential equations. For the more general functional equation these questions require investigation.

Since I wrote this paper Dr R. Bellman has very kindly sent me a manuscript copy of a paper of his (Bellman, 1949) on a similar problem, viz. that in which $\Lambda(y)$ contains $y^{(n)}(x + b_m)$ and no other $y^{(n)}$. Such an equation can be reduced to a set of n simultaneous equations containing no derivative of order higher than the first, and so he studies in detail only the latter type of equation. When $\Lambda(y)$ contains more than one $y^{(n)}$, this reduction method is not applicable.

2. We take α a fixed positive number and write $\psi(x, w)$ for a continuous function of x and of the variables

$$w_{\mu\nu} \quad (0 < \mu < m, \quad 0 < \nu < n-1)$$

throughout the region R , in which

$$x > 0, \quad |w_{\mu\nu}| < \alpha.$$

We suppose that, in R , whenever

$$\sum_{\mu} \sum_{\nu=0}^{n-1} |w_{\mu\nu}| < W,$$

then

$$|\psi(x, w)| < W\chi(W), \quad (2.1)$$

where $\chi(W)$ is a bounded function of W alone for $W < m(n-1)a$ and such that

$$\int_{0+} \frac{\chi^2(W)}{W} dW \quad (2.2)$$

converges.

In what follows K always denotes a positive number, not always the same at each occurrence, independent of $x, X, \xi, y, z, w, t, \epsilon, \delta$, but possibly depending on some or all of

$$m, n, b_\mu, a_{\mu\nu}, c, d, a. \quad (2.3)$$

The constant implied in the $O(\)$ notation is a K . Both ϵ and δ are positive numbers, to be thought of as small, the choice of δ is subsequent to that of ϵ .

We shall consider the equation

$$\Lambda(y) + \psi(x, y) = 0, \quad (2.4)$$

where Λ is the linear operator defined in (1.2) and $\psi(x, y)$ denotes the result obtained on putting $w_{\mu\nu} = y^{(\nu)}(x + b_\mu)$ in $\psi(x, w)$. We say that y satisfies a δ -set of initial conditions if the usual set of initial conditions is such as to make

$$|y^{(n)}(0)| < \delta, \quad |y^{(n)}(x)| < \delta \quad (0 < x < b_m).$$

By repeated use of (1.1) we see that these imply that

$$|y^{(n)}(x)| < K\delta \quad (0 < x < b_m), \quad (2.5)$$

except possibly for $y^{(n)}(b_m)$.

If $a_{mn} \neq 0$, the zeroes of $\tau(s)$ have their real parts bounded above (see, for example, Lemma 2 (iii) of Wright 1949 a). We denote the least upper bound of these real parts by $-d$, and shall here suppose that $d > 0$.

We shall prove the following

THEOREM.—If $a_{mn} \neq 0$, if $0 < c < d$, and if $\psi(x, y)$ satisfies the above conditions, then, for every $\epsilon > 0$, there exists a $\delta > 0$ such that any solution y of (2.4) satisfying δ -set of initial conditions exists and satisfies

$$|y^{(\nu)}(x)| < \epsilon e^{-cx} \quad (\nu < n) \quad (2.6)$$

for all $x > 0$.

We may clearly always replace ϵ by any smaller number without loss of generality, and we shall always suppose that $\epsilon < \frac{1}{2}a$. In general

the statement "for $\epsilon < K$ " or "for $\delta < K$ " is to be understood, but δ is always chosen subsequently to ϵ , and neither may be chosen to depend on x (or on X or on ξ).

3. The question of existence is fairly trivial. We prove

LEMMA.—For $\epsilon < K$, there is a number $\beta = \beta(\epsilon) > 0$ independent of X such that, if a solution of (2.4) exists and satisfies (2.6) for $0 < x < X$, where $X > b_m$, the solution exists for $0 < x < X + \beta$.

Let us suppose that $X < x < X + b_m - b_{m-1}$. If we put $x - b_m$ for x in (2.4) and write

$$u^{(n)}(x) = \sum_{\mu} a_{\mu n} y^{(n)}(x + b_{\mu} - b_m),$$

we transform (2.4) into

$$u^{(n)}(x) = \psi_1(x, u, y). \quad (3.1)$$

In this equation ψ_1 is a function of x , of the $u^{(v)}(x)$ for which $v < n$ and of the

$$y^{(v)}(x + b_{\mu} - b_m) \quad (\mu < m, v < n). \quad (3.2)$$

Also, by the properties of $\psi(x, y)$, provided each of these u and y functions is suitably bounded, ψ_1 is bounded uniformly with respect to x and X and is a continuous function of x , of the $u^{(v)}(x)$ with $v < n$ and of the functions (3.2).

Since $x + b_{\mu} - b_m < X$ for $\mu < m$, each function of (3.2) has its modulus less than ϵ , is continuous and may be regarded as known. We may then regard (3.1) as a differential equation in the unknown function $u(x)$, in which ψ_1 is a continuous function of x and of the $u^{(v)}(x)$ with $v < n$, which is bounded uniformly with respect to x and X , provided the $u^{(v)}(x)$ are suitably bounded for $v < n$. A standard result from the theory of differential equations (see, for example, Satz 2 of Perron, 1918) then shows that $u^{(v)}(x)$ exists and is continuous for $v < n$, and $X < x < X + \beta < X + b_m - b_{m-1}$, for some positive $\beta = \beta(\epsilon)$ independent of X . By (3.1) the same is true for $u^{(n)}(x)$. Now, by hypothesis, $y^{(n)}(x + b_{\mu} - b_m)$ exists and is integrable for $\mu < m$, and so the same is true for

$$a_{mn} y^{(n)}(x) = u^{(n)}(x) - \sum_{\mu < m} a_{\mu n} y^{(n)}(x + b_{\mu} - b_m),$$

provided that $X < x < X + \beta$. Since $a_{mn} \neq 0$, this is the result of the lemma.

Now, for $v < n$ and $0 < x < b_m$, $y^{(v)}(x)$ is given by (1.1) and the initial conditions independently of the value of $y^{(n)}(b_m)$. The latter is then defined by (2.4) with $x=0$. Hence, for $\delta = \delta(\epsilon)$ sufficiently small,

the hypothesis of the lemma is satisfied for $X=b_m$ by any solution satisfying a δ -set of initial conditions. Hence we may apply the lemma in successive steps.

From this we see that, if our solution ceases to exist at any value of x , (2.6) must have been violated for some smaller positive value of x and for at least one value of $\nu < n$, say $\nu = \lambda$. But $y^{(\nu)}(x)$ ($\nu < n$) is continuous by (1.1), and so, if our theorem is false, there must be a number $\xi > b_m$ and a $\lambda < n$, such that

$$|y^{(\lambda)}(\xi)| = \epsilon e^{-\alpha \xi} \quad (3.3)$$

and

$$|y^{(\nu)}(x)| < \epsilon e^{-\alpha x} \quad (0 < x < \xi) \quad (3.4)$$

for all $\nu < n$.

Our theorem does not assert that y is unique for all $x > 0$. It is enough to ensure uniqueness if $\psi(x, w)$ satisfies a Lipschitz condition of order 1 with respect to every w_m .

4. We now assume (3.3) and (3.4) true for a particular ξ and show that they lead to a contradiction. This will complete the proof of the theorem.

We write

$$v(x) = \begin{cases} -\psi(x, y) & (0 < x < \xi - b_m) \\ 0 & (x > \xi - b_m) \end{cases}$$

and consider the linear equation

$$\Lambda(s) = v(x) \quad (4.1)$$

in the unknown function s . If we suppose s to satisfy the initial conditions,

$$s^{(\nu)}(0) = y^{(\nu)}(0) \quad (\nu < n), \quad s^{(n)}(x) = y^{(n)}(x) \quad (0 < x < b_m),$$

it follows from Theorem 1 of Wright (1949*a*) that $s(x)$ is uniquely defined for all $x > 0$. If we put $s(x) = y(x)$ for $0 < x < \xi$, (4.1) is satisfied for $0 < x < \xi - b_m$ by (2.4). Hence

$$s^{(\nu)}(x) = y^{(\nu)}(x) \quad (0 < x < \xi).$$

Next we observe that, since $v(x) = 0$ for $x > \xi$,

$$\int_0^\infty v^2(x) e^{-2\sigma x} dx$$

is absolutely convergent for all σ , and so, by Theorem 1 of Wright (1948),

$$\int_0^\infty |s^{(\nu)}(x) e^{-\sigma x}|^2 dx$$

is convergent for every $\sigma > -d$ and every ν . But

$$\left(\int_0^x |z^{(\nu)}| e^{cx} dx \right)^2 < \int_0^x |z^{(\nu)}|^2 e^{(c+d)x} dx \int_0^x e^{-(d-c)x} dx$$

by Schwarz's inequality, and so

$$\int_0^\infty |z^{(\nu)}| e^{cx} dx$$

is convergent for every ν .

For the rest of this paper we take the real part of s equal to $-c$ and write $s = -c + it$. If

$$Z_\nu = Z_\nu(s) = \int_0^\infty z^{(\nu)} e^{-sx} dx,$$

the integral is absolutely convergent and, for $x > 0$,

$$z^{(\lambda)}(x) = \frac{e^{-cx}}{2\pi} \int_{-\infty}^{\infty} Z_\lambda(s) e^{isx} ds \quad (4.2)$$

by Theorem 7.3 of Widder (1941), since $z^{(\lambda)}(x)$ is an integral and so continuous and of bounded variation.

If we multiply (4.1) by e^{-sx} and integrate from 0 to $+\infty$, we have

$$\sum_\mu \sum_\nu a_{\mu\nu} e^{b_\mu s} Z_\nu(s) = V(s) + H(s), \quad (4.3)$$

where

$$V(s) = \int_0^\infty v(x) e^{-sx} dx, \quad H(s) = \sum_\mu \sum_\nu a_{\mu\nu} H_{\mu\nu}(s),$$

$$H_{\mu\nu}(s) = e^{b_\mu s} \int_0^{b_\mu} z^{(\nu)}(x) e^{-sx} dx = \int_0^{b_\mu} z^{(\nu)}(b_\mu - x) e^{sx} dx.$$

Now

$$\int_0^\infty z^{(\nu)}(x) e^{-sx} dx = \left[z^{(\nu-1)}(x) e^{-sx} \right]_0^\infty + s \int_0^\infty z^{(\nu-1)}(x) e^{-sx} dx.$$

Since both integrals are absolutely convergent, $z^{(\nu-1)}(x) e^{-sx} \rightarrow 0$ as $x \rightarrow \infty$, and so

$$Z_\nu = s Z_{\nu-1} - z^{(\nu-1)}(0).$$

Using this repeatedly we have, if $\nu > \lambda$,

$$\begin{aligned} s^\lambda Z_\nu &= s^\nu Z_\lambda - \sum_{\omega=\lambda}^{\nu-1} s^{\lambda+\nu-\omega-1} z^{(\omega)}(0) \\ &= s^\nu Z_\lambda - s^{\nu-1} z^{(\lambda)}(0) + O(\delta s^{\nu-2}). \end{aligned} \quad (4.4)$$

On the other hand, if $\nu < \lambda$,

$$\begin{aligned} s^\lambda Z_\nu &= s^\nu Z_\lambda + \sum_{\omega=\nu}^{\lambda-1} s^{\lambda+\nu-\omega-1} z^{(\omega)}(0) \\ &= s^\nu Z_\lambda - s^{\nu-1} z^{(\lambda)}(0) + O(\delta s^{\lambda-1}). \end{aligned} \quad (4.5)$$

Using (4.4) and (4.5) in (4.3), we have

$$\tau(s)Z_\lambda = s^{-1}\tau(s)s^{(\lambda)}(0) + s^\lambda V(s) + s^\lambda H(s) + O(\delta s^{n-\lambda}). \quad (4.6)$$

It is well known that, since $c > 0$,

$$\int_{-c-i\infty}^{-c+i\infty} \frac{e^{isz}}{s} ds = 0 \quad (4.7)$$

when $x > 0$. Again, every zero of $\tau(s)$ has its real part less than or equal to $-d < -c$. Hence by Lemma 2 (iv) of Wright (1949 a), since $s = -c + it$,

$$|\tau(s)| > K |s|^n. \quad (4.8)$$

Using (4.6)–(4.8) in (4.2), we have

$$|s^{(\lambda)}(x)e^{sx}| < K \int_{-\infty}^{\infty} \left(\frac{|V(s) + H(s)|}{|s|} + \frac{\delta}{|s|^2} \right) dt. \quad (4.9)$$

Now

$$\int_{-\infty}^{\infty} \frac{dt}{|s|^2} = \int_{-\infty}^{\infty} \frac{dt}{c^2 + t^2} = K,$$

and so, by Schwarz's inequality,

$$\begin{aligned} \left(\int_{-\infty}^{\infty} \frac{|V(s) + H(s)|}{|s|} dt \right)^2 &< K \int_{-\infty}^{\infty} |V(s) + H(s)|^2 dt \\ &< K \int_{-\infty}^{\infty} \left(|V(s)|^2 + \sum_{\mu} \sum_{\nu} |H_{\mu\nu}(s)|^2 \right) dt. \end{aligned}$$

By Parseval's Theorem and (2.5),

$$\int_{-\infty}^{\infty} |H_{\mu\nu}(s)|^2 dt = \int_0^{\infty} |z^{(\nu)}(b_\mu - x)|^2 e^{2cx} dx < K\delta^2,$$

and, by (2.1) and (3.4),

$$\begin{aligned} \int_{-\infty}^{\infty} |V(s)|^2 dt &= \int_0^{\infty} |v(x)|^2 e^{2cx} dx = \int_0^{t-b_m} |\psi(x, y)|^2 e^{2cx} dx \\ &< K\epsilon^2 \int_0^{\infty} \chi^2(K\epsilon e^{-2cx}) dx \\ &< K\epsilon^2 \int_0^{K\epsilon} \frac{\chi^2(W)}{W} dW = \epsilon^2 A^2(\epsilon), \end{aligned}$$

where $A(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$, by (2.2).

Hence, by (4.9),

$$|s^{(A)}(x)e^{\alpha x}| < K_1\delta + K_2\epsilon A(\epsilon).$$

We now choose ϵ small enough to make $K_2A(\epsilon) < \frac{1}{2}$ and then δ small enough to make $K_1\delta < \frac{1}{2}\epsilon$. Hence, when $x = \xi$,

$$|s^{(A)}(\xi)| < \frac{1}{2}\epsilon e^{-\alpha\xi},$$

which contradicts (3.3).

REFERENCES TO LITERATURE

- BELLMAN, R., 1947. "On the boundedness of solutions of non-linear differential and difference equations", *Trans. Amer. Math. Soc.*, LXII, 357-386.
- , 1949. "On the existence and boundedness of solutions of non-linear differential-difference equations", *Annals of Math.*, L, 347-355.
- CESARI, L., 1939. "Sulla stabilità delle soluzioni delle equazioni differenziali lineari", *Ann. Scuola norm. super. Pisa*, VIII, 131-148.
- LEVINSON, N., 1946. "The asymptotic behaviour of a system of linear differential equations", *Amer. Journ. Math.*, LXVIII, 1-6.
- LIAPOUNOFF, A., 1907. "Problème générale de la stabilité du mouvement", *Ann. Fac. Sci. Univ. Toulouse* (2), IX, 203-475.
- PERRON, O., 1918. "Ein neuer Existenzbeweis fuer die Integrale eines Systems gewoehnlicher Differentialgleichungen", *Math. Ann.*, LXXVIII, 378-384.
- , 1928. "Ueber Stabilitaet und asymptotisches Verhalten der Loesungen eines Systems endlicher Differenzengleichungen", *Journ. fuer Math.*, CLXI, 41-64.
- , 1929. "Ueber Stabilitaet und asymptotisches Verhalten der Integrale von Differentialgleichungssystemen", *Math. Zeits.*, XXIX, 129-160.
- PITT, H. R., 1944. "On a class of integro-differential equations", *Proc. Camb. Phil. Soc.*, XL, 199-211.
- , 1947. "On a class of integro-differential equations. II", *Proc. Camb. Phil. Soc.*, XLIII, 153-163.
- POINCARÉ, H., 1892. *Les méthodes nouvelles de la mécanique céleste*, Paris.
- WEYL, H., 1946. "Comment on the preceding paper", *Amer. Journ. Math.*, LXVIII, 7-12.
- WIDDER, D. V., 1941. *The Laplace transform*, Princeton, Chapter II.
- WRIGHT, E. M., 1946. "The non-linear difference-differential equation", *Quart. Journ. Math.*, XVII, 245-252.
- , 1948. "The linear difference-differential equation with asymptotically constant coefficients", *Amer. Journ. Math.*, LXX, 221-238.
- , 1949 a. "The linear difference-differential equation with constant coefficients", *Proc. Roy. Soc. Edin.*, A, LXII, 387-393.
- , 1949 b. "Perturbed functional equations", *Quart. Journ. Math.*, XX, 155-165.

III.—The Simplest Form of Second-Order Linear Differential Equation, with Periodic Coefficient, having Finite Singularities.* By Enzo Cambi. *Communicated by Professor A. C. AITKEN, F.R.S. (With Two Text-figures.)*

(MS. received May 11, 1948. Revised MS. received April 4, 1949.
Read January 10, 1949)

SYNOPSIS

A differential equation of the second order, arising in problems of disturbed oscillation, such as occur in frequency modulation, is considered. The nature of its solutions is examined by the method of continued fractions. The cases in which the solutions are periodic, and the regions of stability and instability (lability), are determined according to the values taken by the two parameters involved.

I. INTRODUCTION

1. WHEN the essential parameters of an oscillating system of any nature are periodic functions of the time, the performance of the system is defined by a differential equation, which can be regarded as derived from the harmonic equation

$$\frac{d^2y}{dx^2} + \omega^2 y = 0$$

by replacing the constant ω^2 (depending on the values of the parameters when they are supposed fixed) by a periodic function of the time-variable x .

Depending on the nature of this periodic function two general classes of equations can be defined, according as the function is an integral one, or has finite singularities in the complex x -plane.

The distribution of the singularities of the coefficient of y determines that of the singularities of the solution of the equation. The simplest cases of the above classes are respectively when the periodic coefficient reduces to a binomial expression, of degree 1 or -1 , for instance

$$\frac{d^2y}{dx^2} + p^2(1 - 2\gamma \cos x)y = 0 \quad (1)$$

or

$$\frac{d^2y}{dx^2} + \frac{p^2}{1 + 2\gamma \cos x}y = 0. \quad (2)$$

* The author is greatly indebted to Dr J. C. P. Miller for having revised the English text and for having made many helpful suggestions and remarks.

Equation (1) is the well-known Mathieu equation; equation (2) arises in many problems of *disturbed* oscillations, such as in frequency-modulation, where it gives an exact description of the performance of the system, in the non-dissipative case, that is.

2. It is evident that equations (1) and (2) are equivalent, when γ is so small that its square is negligible.

The Mathieu equation can be regarded as the simplest case of Hill's equation, where the coefficient of y may be expressed in the general form of a Fourier series valid in a strip of the complex plane enclosing the real axis. It is therefore of importance as an approximating equation, providing approximate solutions of a Hill's equation.

The nature of equation (2) substantially depends on the magnitude of the parameter 2γ ; when $|2\gamma| > 1$ the singularities of the equation are real and the coefficient itself cannot be expanded in a Fourier series, so that (2) is not even a Hill's equation in this case.

On the contrary, when $|2\gamma| < 1$ the singularities are complex, and the coefficient is analytic in a strip of the x -plane, bounded by the straight lines containing the points given by $\cos x = \pm 1/2\gamma$, where it can be expanded in a Fourier series. In this case the equation could also be solved by Hill's general method; but unless $|\gamma|$ is very small this method leads to complicated expansions, difficult to interpret physically.

The main purpose of the present paper is to expound a method more convenient than Hill's method, applicable in the case $2\gamma < 1$. In physical applications this case is often the only significant case; for example in frequency-modulation, when 2γ can be regarded as a relative modulation.

The convenience of the method set out below is so remarkable when γ is small that it is thought that equation (2) may conveniently replace the Mathieu equation as an approximating form for Hill's equation.

The cases $|2\gamma| = 1$ and $|2\gamma| > 1$ are discussed separately, since, in these cases, the equation has quite a different nature from that which holds when $|2\gamma| < 1$. Owing to the comparative unimportance of these cases in physical applications, they are treated below in less detail.

II. SOLUTION OF THE SINGULAR EQUATION (2) WHEN $2\gamma < 1$

3. We write equation (2) in the form

$$(1 + 2\gamma \cos x) \frac{d^2 y}{dx^2} + p^2 y = 0, \quad (2')$$

where we regard 2γ and p^2 as real positive parameters, with $2\gamma < 1$. In this and in later sections we regard γ as positive, no loss of generality being incurred thereby.

The coefficient of d^2y/dx^2 being periodic, Floquet's theory ensures the existence of solutions that may be represented in the form of the product of an exponential term by a periodic function, expressible as a Fourier series valid in the strip above defined,

$$y = e^{i u_0 x} \sum_{-\infty}^{\infty} B_n e^{i n x}. \quad (3)$$

We prefer here the form $e^{i u_0 x}$ to the more usual $e^{u_0 x}$. If u_0 is real, $e^{i u_0 x}$ represents a (central) trigonometric component of period $u_0/2\pi$.

In this expression u_0 is evidently indeterminate to the extent of an additive integer. Further, since the left-hand side of equation (2') is even in x , the expression for y obtained by replacing x by $-x$ is also an integral of the equation, obtained formally from (3) by changing the sign of u_0 . Hence: *Any equation giving the condition to be satisfied by u_0 must at the same time be satisfied by a number of the form $\pm(u_0 \pm n)$.*

By substituting for y , in (2'), a series of the form (3), where u is written for u_0 , and by equating to zero the coefficient of every component $\exp[i(u \pm n)x]$, we obtain an infinite system of homogeneous linear equations satisfied by the B_n . The general equation of the system may be written as

$$\gamma A_{n+1} + G(u+n)A_n + \gamma A_{n-1} = 0, \quad n = \dots, -2, -1, 0, 1, 2, \dots, \quad (4)$$

where *

$$A_n = (u+n)^2 B_n / p^2, \quad G(u) = 1 - \frac{p^2}{u^2}.$$

In considering the above general equation it is logical to regard A_n as a function of $u+n$, let us say $A_n = V(u+n)$; we then find that all equations (4) for which the suffixes n are all positive (we shall call these *positive* equations) are satisfied for any value of u , provided that the function V satisfies the linear difference equation † of second order

$$\gamma V(u+1) + G(u)V(u) + \gamma V(u-1) = 0. \quad (5)$$

Similarly the *negative* equations (4), $n < 0$, are satisfied if A_{-n} is

* If u is such that the series $e^{i u x} \sum B_n e^{i n x}$ is an integral of (2'), then the series $e^{i u x} \sum A_n e^{i n x}$ is proportional to d^2y/dx^2 , that is, to $y/(1 + 2\gamma \cos x)$.

† For information concerning difference equations, the treatise by N. E. Nörlund, *Vorlesungen über Differenzenrechnung*, Berlin, 1924, may be consulted.

replaced by $V(-u+n)$, where the function V satisfies *the same difference equation as above*; this follows since $G(u)$ is an even function* of u .

All the integrals of a difference equation of the second order, such as (5), are known as soon as two linearly independent integrals constituting a fundamental system are known. If we divide (5) by $V(u)$ and solve by recurrence for $V(u)/V(u+1)$ or for $V(u)/V(u-1)$, we obtain two independent expressions for these ratios, given in the form of continued fractions as follows:—

$$\frac{V(u)}{V(u+1)} = \frac{-\gamma}{G(u)} - \frac{\gamma^2}{G(u-1)} - \frac{\gamma^2}{G(u-2)} - \dots,$$

$$\frac{V(u)}{V(u-1)} = \frac{1}{G(u)} - \frac{\gamma^2}{G(u+1)} - \frac{\gamma^2}{G(u+2)} - \dots$$

Since $G(u)$ is even, we can express both ratios in terms of a single function of u defined by the continued fraction

$$v(u) = \frac{1}{G(u)} - \frac{\gamma^2}{G(u+1)} - \frac{\gamma^2}{G(u+2)} - \dots,$$

which converges for all values of u and p , provided that $2\gamma < 1$. Two different and independent expressions for the ratio of two consecutive values of the function V , both satisfying the difference equation, are then

$$\frac{V(u)}{V(u+1)} = -\gamma v(-u); \quad \frac{V(u)}{V(u-1)} = -\gamma v(u). \quad (6)$$

Assuming an arbitrary initial determination of $V(u)$ in a unit interval of real values of u , we may, by repeated multiplication by one or other of the above expressions for the ratio, define two independent solutions of the difference equation throughout the whole real u -axis.

The resulting functions are not as a rule analytic, but may, for example, have discontinuities at all points which are congruent with the limits of the initial interval. Analytic solutions of (5) may actually be obtained, but their determination is unnecessary for our purpose, since only the values at congruent arguments are required.

4. The continued fraction $-\gamma v(u)$ tends to the value

$$\frac{-\gamma}{1-\gamma} - \frac{\gamma^2}{1-\gamma} - \frac{\gamma^2}{1-\gamma} - \dots = -(1 - \sqrt{1-4\gamma^2})/2\gamma,$$

* If A_0 is regarded as belonging to the *positive* system, for $n=0$, it is expressed as $V(u)$; if it is regarded as a term of the *negative* system A_{-n} , it is given by $V(-u)$. The discrepancy is only apparent, however, since the function V is undetermined to the extent of an arbitrary factor, and A_0 itself is an *arbitrary* constant of the differential equation.

the modulus of which is smaller than unity * (if $2\gamma < 1$) when u tends to infinity through positive values.

Assuming arbitrarily $V(u) = A_0$, and evaluating $V(u+1)$, $V(u+2)$, . . . by means of the second of the ratios (6), thus

$$A_1 = V(u+1) = -\gamma v(u+1)A_0; \quad A_2 = -\gamma v(u+2)A_1 = \gamma^2 v(u+2)v(u+1)A_0,$$

and so on, we obtain a succession of numbers A_n , converging to zero as n tends to infinity, and satisfying the positive set of equations (4). Similarly, assuming $V(-u) = A_0$, and evaluating

$$A_{-1} = V(-u+1) = -\gamma v(-u+1)A_0; \quad A_{-2} = V(-u+2) = \gamma^2 v(-u+2)v(-u+1)A_0,$$

and so on, we obtain a set of values satisfying the *negative* equations of the system, and also converging to zero as n tends to infinity.

The equations of the linear system belonging to values of n respectively positive or negative are thus satisfied *independently* by the above expressions, *for any value of u* . The possible values which can actually be assumed by u are then defined by the central equation, for $n=0$, which has not yet been considered, and which we may write as

$$\gamma \frac{A_1}{A_0} + G(u) + \gamma \frac{A_{-1}}{A_0} = 0,$$

that is,

$$-\gamma^2 v(-u+1) + G(u) - \gamma^2 v(u+1) = 0. \quad (7)$$

5. Equation (7) replaces Hill's determinantal equation, to which it may be shown to be equivalent. We shall refer to it below as the "resonance equation", since, if u_0 is real, it gives the natural frequencies of the oscillating system. If u_0 is a root of (7) and the terms A are computed by the above recurrences, with the assumption that $A_0 = 1$, the series

$$e^{iu_0 x} \sum_{-\infty}^{\infty} A_n e^{in x}$$

represents the second derivative $d^2 y/dx^2$ of an integral of (2'); that is, in virtue of the equation itself, and integral of (2') divided by $1 + 2\gamma \cos x$.

A series representing the integral itself may be written as

$$e^{iu_0 x} \sum_{-\infty}^{\infty} B_n e^{in x},$$

where $B_n = u_0^2 A_n / (u_0 + n)^2$, so that we have also $B_0 = 1$.

* If $2\gamma > 1$ the continued fractions diverge in any case, and the difference equation cannot be solved by means of continued fractions (Nörlund, 1924, p. 441).

III. DISCUSSION OF THE SOLUTION

6. In virtue of its definition, the function $v(u)$ satisfies the recurrence relation

$$\gamma^2 v(u+1) = G(u) - \frac{1}{v(u)}, \quad (8)$$

from which, by recurrence, any value of the function may be determined as soon as it is known in a given unit interval of values of the argument.

In particular, the functions v appearing in the general coefficient

$$A_{\pm n} = (-\gamma)^n v(\pm u_0 + 1) v(\pm u_0 + 2) \dots v(\pm u_0 + n),$$

can be deduced from the initial values $v(1 \pm u_0)$, the computation of the continued fraction being required only for these two values.

In view of the recurrence formulæ (8), the resonance equation (7) can be written

$$\frac{1}{\gamma v(u)} - \gamma v(-u+1) = 0 \quad \text{or} \quad \gamma^2 v(u) v(1-u) = 1. \quad (7')$$

It is easily seen, again by (8), and by the fact that $G(u)$ is even, that if u_0 is a root, then any number of the form $\pm(u_0 \pm n)$, where n is an integer, also satisfies the equation, in conformity with the general conclusion regarding the exponential term of Floquet's solution.

The roots of (7'), however, cannot all be determined with equal ease, since the roots that are relatively distant from p and $-p$ respectively are extremely difficult to locate, and may also prove refractory to numerical tabulation, even if the tabular interval is very small. The "lateral" roots owe their existence, in fact, to the existence of odd poles of the expressions on the left side of (7'), the sharpness of which increases rapidly (fig. 1) with increasing distance from $|u| = |p|$.

The root u_0 , which is close to $u=p$ and tends* to p when $\gamma \rightarrow 0$, is on the contrary determined very easily, since the left side of the resonance equation, written in the first of the forms (7'), varies almost linearly with u in the vicinity of $u=p$; all the other roots are then automatically known also. The determination of the "characteristic exponents", which is the crucial point in the resolution of any differential equation with periodic coefficients, is thus formulated and solved, in the case of the present equation, with exceptional simplicity.

* In the case $\gamma=0$, the resonance equation is reduced to $G(u)=0$, that is, $u^2=p^2$. The differential equation is then simply a harmonic equation the general integral of which is $C_1 e^{i p x} + C_2 e^{-i p x}$.

All the roots of (7') that differ by integers from one another give rise to the same set of values of the amplitudes A_n (in relative magnitude), that is, to the same integral of the differential equation. Thus, as is to be expected, only two independent integrals of the equation itself can be obtained, depending on the form $(n + u_0$ or $n - u_0)$ of the chosen value of u , where u_0 is the "central" positive root, namely the one tending to p when $\gamma \rightarrow 0$.

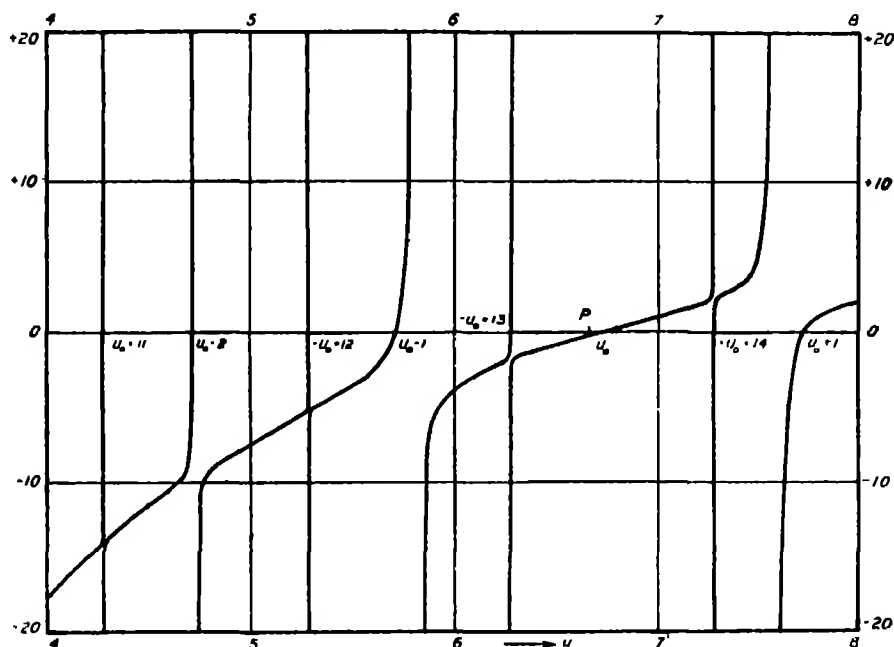


FIG. 1.—Behaviour of the expression $1/[\gamma v(u)] - \gamma v(-u + 1)$ around $u = p$.
[$1/p = 0.15$; $\gamma = 0.1$.]

7. The ratio of two consecutive coefficients, A_{n+1}/A_n , or A_{-n-1}/A_{-n} , of the Fourier series $\sum A_n e^{inx}$, that is, the number $-\gamma v(n + 1 \pm u_0)$, tends to $-(1 - \sqrt{1 - 4\gamma^2})/2\gamma$ as $n \rightarrow \infty$. This is smaller than unity in absolute value, since $2\gamma < 1$, and the series converges absolutely for all values of x such that

$$|e^{\pm ix}| < |(1 - \sqrt{1 - 4\gamma^2})/2\gamma| = |(1 + \sqrt{1 - 4\gamma^2})/2\gamma|.$$

These values of x are contained in the strip of the complex plane limited by the straight lines

$$|I(x)| = |\log [(1 + \sqrt{1 - 4\gamma^2})/2\gamma]|,$$

which contain the singularities given by $2\gamma \cos x = -1$.

IV. APPROXIMATE EXPRESSIONS, VALID IN THE CASE OF SMALL γ

8. The first step of the solution is the determination of one root of the resonance equation. The determination from equation (7) of the exact value of the root u_0 lying near p is not difficult, unless 2γ is extremely close to unity. An approximate expression, valid when γ is small, on which further approximations may be based, is easily obtained by replacing the continued fraction $v(u)$ by the first convergent $1/G(u)$; this can be done with a relative error of the order of γ^2 . It must be supposed that $p \neq \frac{1}{2}$. The left side of (7), when $u_0 = p$, has the value

$$-\gamma^2 v(p+1) - \gamma^2 v(-p+1) \simeq -2\gamma^2(3p^2-1)/(4p^2-1),$$

and its first derivative, to the same order of accuracy, is $2/p$; so that an approximate value of the root is given by

$$u_0 \simeq p\{1 + \gamma^2(3p^2-1)/(4p^2-1)\}$$

with an error of the order of γ^4 . Correspondingly,

$$A_{\pm n}/A_{\pm(n-1)} \simeq -\gamma v(n \pm u_0) \simeq -\gamma(n \pm u_0)^2 / \{(n \pm u_0)^2 - p^2\}; \quad (9)$$

the value of u_0 to be introduced in this expression can be deduced, for instance, from the above approximate formula.

If the difference $u_0 - p$ may be neglected, we can assume for the ratio the simple expression

$$A_{\pm n}/A_{\pm(n-1)} \simeq -\gamma(n \pm p)^2 / n(n \pm 2p). \quad (10)$$

The corresponding expressions for the coefficients A , assuming $A_0 = 1$, are respectively

$$\begin{aligned} A_{\pm n} &\simeq (-\gamma)^n \frac{(n \pm u_0)!^2}{(\pm u_0)^2} \frac{(p \pm u_0)(-p \pm u_0)}{(n \pm p \pm u_0)!(n - p \pm u_0)!} \\ &\simeq (-\gamma)^n \frac{(n \pm p)!^2}{(\pm p)^2} \frac{(\pm 2p)}{n!(n \pm 2p)!} \end{aligned}$$

with the usual extension of the meaning of the symbol $x!$ to non-integral values of the argument.

V. A NUMERICAL EXAMPLE

9. The performance of the simplest frequency-modulating circuit, that is, of a variable-capacitance (or variable-inductance) resonant circuit, is given, in the non-dissipative case, by equation (2), where p is the ratio of the (static) resonance frequency to the modulating frequency, and 2γ is the relative modulation of the variable element. In the present example

we assume $p=20/3$, $2\gamma=0.2$. (Such values, we may observe, are unusual in the technical applications, where p is usually very large and γ very small.)

The approximate formula of § 8 then gives for the central resonance the value $u_0=6.71657$, to an accuracy of the order of $\gamma^4=0.0001$.

If we compute the expression on the left of the resonance equation

$$P(u_0) = -\gamma^2 v(1-u_0) + G(u_0) - \gamma^2 v(1+u_0) = 0$$

for this value of u_0 , we obtain -0.0004182359 . Since this is negative, and since the slope of the corresponding curve is positive and of the order $2/p=0.3$ but slightly greater than 0.3 , the actual root is larger than the above approximation, but smaller than that which could be found by assuming a slope of 0.3 for the curve. We thus locate the root between 6.71657 and, say, 6.718 .

We find, in fact, that $P(6.718) = \pm 0.0001131942$. Linear interpolation then gives an improved value of u_0 , namely $u_0' = 6.7176954111$. Similarly, $P(u_0') = \pm 0.000000021538$, and a further linear interpolation, using this and the former $P(6.718)$, gives $u_0'' = 6.717695353134$, a value which satisfies the resonance equation to within 10^{-12} .

The values of the v -functions used in subsequent computations are

$$\begin{aligned} -\gamma^2 v(1+u_0'') &= -0.043754163443 \\ -\gamma^2 v(1-u_0'') &= 0.028619545755 \end{aligned}$$

and

$$G(u_0'') = 0.015134617688$$

The terms $v(n \pm u_0)$ appearing in the expressions for the amplitudes A_n can now be deduced by means of (8).

Remark.—When computing with the recurrence formula (8), which gives $v(n+1)$ in terms of $v(n)$, a loss of accuracy is unavoidable; and the same happens in deducing B_n from A_n (§§ 3, 5) when n is negative. In order to give all the coefficients to the same degree of accuracy, obvious artifices are used, such as (i) direct computation of the highest $v(n+u_0)$ by the continued fraction—so as to apply the recurrence backwards; (ii) independent computation of B_{-n} (from an obvious recurrent expression) and deduction of A_{-n} from it.

Since $B_0=A_0=1$, if we assume the (B) series for y , the derivative d^2y/dx^2 will be given by the (A) series multiplied by $-u_0^2$; the differential equation then becomes

$$-(1 + \gamma e^{ix} + \gamma e^{-ix}) \frac{u_0^2}{p^2} \sum A_n e^{i(n+u_0)x} + \sum B_n e^{i(n+u_0)x} = 0;$$

and this may be used as a numerical check.

We give the values of the coefficients, as deduced from the above formulæ, as follows:—

n	B_n	$-\frac{\mu_0^2}{\rho^2} A_n$	$-\gamma \frac{\mu_0^2}{\rho^2} A_{n-1}$	$-\gamma \frac{\mu_0^2}{\rho^2} A_{n+1}$	Sum
-6	190	-2	0	-187	+0*
-5	28242	-1875	0	-26367	0
-4	15 86654	-2 63674	-187	-13 22794	-1*
-3	425 36519	-132 27936	-26367	-292 82216	0
-2	5847 38436	-2928 22163	-13 22794	-2905 93479	-0*
-1	+0.39505 84199	-0.29059 34788	-292 82216	-0.10153 67194	+1
0	+1.00000 00000	-1.01536 71943	-2905 93479	4442 65422	0
1	-0.33150 09151	+0.44426 54217	-0.10153 67194	-1122 77872	0
2	6566 11189	-0.11227 78722	4442 65422	219 02111	0
3	-1030 80589	2190 21108	-1122 77872	-36 62646	+1
4	141 71297	-366 26460	219 02111	5 53052	0
5	-17 90191	55 30524	-36 62646	-77687	0
6	2 13477	-7 76874	5 53052	10344	-1
7	-24431	1 03440	-77687	-1322	+0*
8	2712	-13220	10344	164	+0*
9	-294	1636	-1322	-20	0
10	31	-197	164	2	0
11	-3	23	-20	0	0
12	0	-3	2	0	-1

Decimal points and ciphers are omitted except in the largest entries. The entries in the "Sum" column are in the tenth decimal place; an asterisk indicates that a 5, enforced, follows in the eleventh place.

A comparison between the accurate values and those which can be obtained by using the approximate asymptotic formulæ (9) and (10) may be useful; the comparison is shown in the following table, where the accurate values of A_n (Column I)* are compared with those of the formula (9) (taken with the true value of μ_0) and of the formula (10) (Columns II and III):—

n	I	II	III
-6	2	2	I
-5	1847	1785	1378
-4	2 59683	2 50998	2 06644
-3	130 27736	125 93840	108 48806
-2	2883 90412	2790 36452	2501 50150
-1	0.28619 54576	0.27817 32090	0.26036 03604
0	1.00000 00000	1.00000 00000	1.00000 00000
1	-0.43754 16344	-0.39397 62919	-0.41007 75194
2	0.11057 85895	9489 03916	0.10043 92765
3	-2157 06307	-1792 55699	-1915 40061
4	360 72133	292 38273	314 32314
5	-54 46822	-43 23224	-46 67222
6	7 65116	5 96134	6 45543
7	-1 01875	-78047	-84712
8	13020	9819	10677
9	-1611	-1197	-1304
10	194	142	155
11	-23	-16	-18
12	3	2	2

* In many physical problems the coefficients A_n are of greater importance than the B_n , which in any case are readily deduced from them (§ 16).

If we note that the assumed value for γ (namely 0.1) can actually be regarded as very high for most physical applications, we readily conclude that the approximate formulæ following (9) and (10) (the latter being extremely simple for computation) are very convenient for a preliminary survey of the approximate distribution of the components.

VI. STABILITY OF THE SOLUTIONS

10. Any differential equation with periodic coefficients always has regions of instability (*lability*); that is, values of the parameters exist, for which one or both of the independent integrals increase or decrease indefinitely in amplitude as the independent variable increases. It is obvious that, so long as u_0 is real, the absolute value of an expression such as $\exp(\pm i u_0 x) \sum B_n \exp(\pm i n x)$ remains unchanged when x increases by 2π . If u_0 is complex, on the other hand, each of the two integrals is then multiplied by a constant, one less and one greater than unity, when x increases by 2π . More precisely, Floquet's theory states that two independent integrals of the second-order equation may be obtained, each satisfying a relation of the form $F(x+2\pi) = \rho F(x)$, where ρ is a complex constant; and that the two possible values of ρ are given by a quadratic equation, depending only on the constants of the original equation.

In the present case, the two possible values of ρ are known in the form $\exp(\pm 2\pi i u_0)$, so that the quadratic equation is written at once as

$$\rho^2 - 2\rho \cos 2\pi u_0 + 1 = 0,$$

having the discriminant $4D^2 = -\sin^2 2\pi u_0$.

So long as u_0 is real, we know that ρ is complex with modulus unity, and the solutions are both stable; ρ can pass to values having different modulus only at points which make the discriminant vanish, that is, values which make $u_0 = n$ or $n + \frac{1}{2}$, where n is any integer.

11. In the (ρ, γ) plane, therefore, only the curves given by pairs of values of ρ and γ making u_0 an integer or the half of an odd integer can be boundaries of the regions of lability, where these exist. The equations of these curves are obtained at once by noting that the resonance equation must be satisfied by $u = n$ or $u = n + \frac{1}{2}$ respectively; that is, owing to the congruence of the roots, by $u = 0$ or $u = \frac{1}{2}$ respectively.

The conditions are obviously $1/v(1) = 0$ or $\gamma^2 v^2(\frac{1}{2}) = 1$; that is,

$$1 - \rho^2 - \frac{\gamma^2}{1 - \frac{\rho^2}{4}} - \frac{\gamma^2}{1 - \frac{\rho^2}{9}} - \dots = 0; \quad (11)$$

$$1 \pm \gamma - 4p^2 - \frac{\gamma^2}{1 - \frac{4p^2}{9}} - \frac{\gamma^2}{1 - \frac{4p^2}{25}} - \dots = 0. \quad (12)$$

In the case of equation (12), two families of curves exist which make $u_0 = n + \frac{1}{2}$, according to the sign preceding the term γ . The two curves corresponding to the same value of u_0 bound, in fact, a region of lability in the (p, γ) plane; they may be found numerically by solving for p^2 by successive approximations with given values of γ .

Equation (11), on the contrary, gives a single family of curves, that is, one curve for each integral value of u_0 ; it will be proved later that there are no regions of lability bounded by curves $u_0 = \text{integer}$. Since u_0 is equal to p at $\gamma = 0$, the curves of families (11) and (12) obviously start at the points $p = n$ or $p = n + \frac{1}{2}$ of the p -axis.

12. Equations (11) and (12) can be solved for p^2 with relative ease, even for values of γ very close to $\frac{1}{2}$. When γ is large, of course, the convergence of the procedure is rather slow, since the (p, γ) curves become very flat near $\gamma = \frac{1}{2}$, where they have a horizontal tangent.* In the numerical case examined above we get the following values for some points of the boundary curves.

VALUES OF p (FOR GIVEN γ) MAKING $u_0 = \frac{1}{2}, 1, 3/2$

	$u_0 = \frac{1}{2}$		$u_0 = 1$	$u_0 = 3/2$	
0.0	0.50000	0.50000	1.00000	1.50000	1.50000
0.1	0.47137	0.52165	0.99326	1.48905	1.48907
0.2	0.43448	0.53659	0.97170	1.45409	1.45428
0.3	0.38463	0.54360	0.93048	1.38673	1.38763
0.4	0.31159	0.53771	0.85374	1.26028	1.26325
0.45	0.25207	0.52348	0.78327	1.14183	1.14789
0.5	0.00000	0.35355	0.35355	(0.35355)	(0.35355)

It is seen at once that the regions of lability comprised between two curves for which u_0 is the same odd multiple of $\frac{1}{2}$, become very narrow as u_0 increases. This is due to the fact that the two curves $u_0 = n + \frac{1}{2}$ have contact of order $2n$ at the point where they cross the p -axis, that is, at the point $p = n + \frac{1}{2}$, $\gamma = 0$, and are all contained in the finite strip $\gamma < \frac{1}{2}$ (fig. 2).

* The computation is actually impracticable when 2γ is just less than 1; a method for solving (11) and (12) with such values of 2γ would thus be of some interest. The problem is, however, of no great practical importance, and from the theoretical point of view, so long as it is desired only to trace the boundary curves, the knowledge of the terminal point at $\gamma = \frac{1}{2}$, $p = \frac{1}{2}\sqrt{2}$ (see § 19) may be regarded as sufficient.

13. When $\gamma = \frac{1}{2}$, the continued fractions $v(u)$ converge or not (Nörlund, 1924, p. 438) according as $1 - 8\rho^2 \gtrless 0$, as long as u_0 remains finite in magnitude. To the right of the point $\rho = \frac{1}{2}\sqrt{2}$ there are therefore no points of boundary curves corresponding to finite values of u_0 .

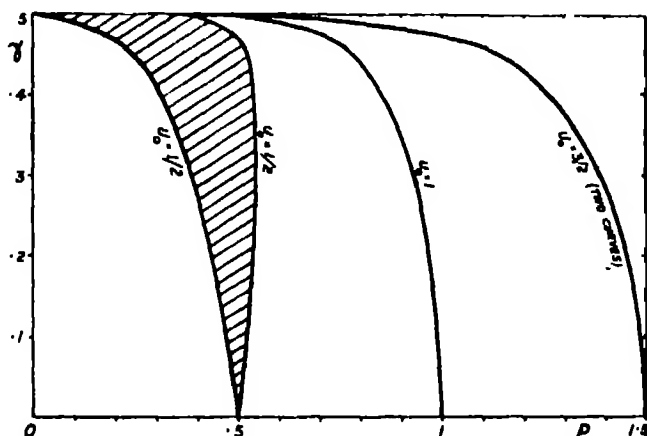


FIG. 2.—The shaded area is a region of liability; the second of such regions runs along $u_0 = 3/2$, and is very narrow.

The point $\rho = \frac{1}{2}\sqrt{2}$, $\gamma = \frac{1}{2}$ is an exceptional point in the representative plane of the parameters. The behaviour of the equation at this point (belonging to the case $2\gamma = 1$) is fully discussed later (§ 18).

VII. CHARACTERISTIC FUNCTIONS OF THE EQUATION

14. When ρ and γ have such values that u_0 is either an integer or an odd multiple of $\frac{1}{2}$, the solutions of the equation are respectively periodic with period 2π , or "half-periodic". In the former case the factor $\rho = \exp(\pm 2\pi i u_0)$, by which each integral is multiplied, has the value ± 1 for an increase 2π in x ; in the latter case the value of the factor is -1 , so that the function has a period 4π , its values in half-periods of length 2π being equal, and opposite in sign.

The integrals are evidently periodic for all *rational* real values of u_0 , the period being then given by 2π times the denominator of u_0 . Since real values of u_0 (making $|\rho| = 1$) are possible only in the regions of stability, the curves (ρ, γ) where periodic solutions exist are contained in these regions and never cross or touch the boundary curves.

Consider now the (A) series, representing the derivative d^2y/dx^2 or $y/(1 + 2\gamma \cos x)$. From the series $\exp(iu_0x) \sum A_n \exp(ikx)$ and

$\exp(-iu_0x)\Sigma A_k \exp(-ikx)$ we deduce two independent solutions * expressed in real form as

$$y_1'' = \sum_{-\infty}^{\infty} A_k \cos(u_0 + k)x, \quad y_2'' = \sum_{-\infty}^{\infty} A_k \sin(u_0 + k)x.$$

If $u_0 = n + \frac{1}{2}$, each of the circular functions appearing in the above series occurs in two terms, so that the half-periodic *eigenfunctions* become

$$K_{n+\frac{1}{2}}(x) = \sum_0^{\infty} (A_{-n+k} + A_{-n-k-1}) \cos(k + \frac{1}{2}x),$$

$$\Sigma_{n+\frac{1}{2}}(x) = \sum_0^{\infty} (A_{-n+k} - A_{-n-k-1}) \sin(k + \frac{1}{2}x).$$

These solutions of the equations can exist when (p, γ) is a point on either of the boundary curves given by $\gamma v(\frac{1}{2}) = \pm 1$. But if we consider the explicit expressions of the constants A in terms of v -functions, we easily see that one of

$$A_{-n+k} \pm A_{-n-k-1} \text{ is zero when } \gamma v(\frac{1}{2}) = \pm 1,$$

so that only one periodic solution exists when (p, γ) is a point on any one of the curves $u_0 = n + \frac{1}{2}$, in conformity with what happens in the case of a Mathieu equation.

15. On the curves $u_0 = n (1/v(1) = 0)$, on the contrary, the equation has two independent periodic solutions of period 2π , namely

$$K_n(x) = \sum_0^{\infty} A_{-n+k} \cos kx \quad \text{and} \quad \Sigma_n(x) = \sum_0^{\infty} A_{-n+k} \sin kx.$$

The solutions are valid along the same curve $u_0 = n$, which is the locus of double eigenvalues.

The possibility of the simultaneous existence of two solutions of period 2π for the same values of the parameters p and γ (making u_0 an integer) may be proved by the same method by which E. L. Ince proved the impossibility of a pair of such solutions, in the case of Mathieu's equation. Assuming that two independent solutions

$$K = \frac{1}{2}a_0 + a_1 \cos x + a_2 \cos 2x + a_3 \cos 3x + \dots$$

$$\Sigma = b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x + \dots$$

can satisfy the differential equation for the same values of p and γ , we easily find, by direct substitution in the equation, the necessary condition

* For the sake of brevity, we speak of the (A) series as solutions of the equation, while the actual solutions, that is, the (B) series, are obtained from the former by division by $(1 + 2\gamma \cos x)$. The reasons for the choice of the (A) series as eigenfunctions are examined in § 16.

$$\begin{vmatrix} a_n & (n+1)^2 a_{n+1} + (n-1)^2 a_{n-1} \\ b_n & (n+1)^2 b_{n+1} + (n-1)^2 b_{n-1} \end{vmatrix} = 0, \quad n=1, 2, 3, \dots \quad (13)$$

which, when solved by recurrence, readily leads to

$$\begin{vmatrix} a_n & a_{n+1} \\ b_n & b_{n+1} \end{vmatrix} = 0,$$

a condition which is obviously satisfied if b_n is proportional to a_n .

If the same procedure is applied to Mathieu's equation, we find the necessary condition

$$\begin{vmatrix} a_n & a_{n+1} + a_{n-1} \\ b_n & b_{n+1} + b_{n-1} \end{vmatrix} = 0, \quad \text{which leads to} \quad \begin{vmatrix} a_n & a_{n+1} \\ b_n & b_{n+1} \end{vmatrix} = a_0 b_1,$$

a condition which is incompatible with the convergence of the Fourier series. The material difference between the two cases is that, in the present one, the first of equations (13) does not contain a_0 , thus differing from the corresponding equation valid in the case of Mathieu's equation.

Once the possibility of the coexistence of the two eigenfunctions has been proved, it is easily verified that the Fourier series set out above (which can be deduced as limiting cases from the general integrals as $\alpha_0 \rightarrow$ an integer) actually satisfy the equation, and constitute one pair of eigenfunctions of integral functions of integral order, corresponding to the even Mathieu functions.

16. In the preceding sections we have considered as characteristic solutions the functions expressed in terms of (A) series, although these actually give the second derivatives of the integrals y of the equation. By substituting the B -coefficients for the corresponding A 's, we might easily write the explicit expressions for the characteristic integrals y .

However, the (A) series actually represent new eigenfunctions of greater functional importance than the corresponding (B) series. It may readily be shown, in fact, that two *different* eigenfunctions of the form (A) with the same coefficient of $\cos x$ in the equation (that is, having two different pairs of parameters, but such that $\gamma_1/p_1^2 = \gamma_2/p_2^2$) are orthogonal to one another over a common period. If we eliminate $\cos x$ between the two identities

$$(1 + 2\gamma_1 \cos x)y_1'' + p_1^2 y_1 = 0, \quad (1 + 2\gamma_2 \cos x)y_2'' + p_2^2 y_2 = 0$$

and integrate term by term, we readily obtain

$$(p_1^2 - p_2^2) \int_{x_0}^{x_1} y_1'' y_2'' dx = \left[y_2 y_1' - y_1 y_2' \right]_{x_0}^{x_1},$$

and the right-hand side vanishes if the interval (x_0, x_1) is a period of both functions.

VIII. THE ALGEBRAIC FORM OF THE EQUATION, AND THE CASE $2\gamma=1$

17. The above trigonometric expansions are applicable only in those cases where $2\gamma < 1$ (cases which are often of more immediate physical interest), since the inequality in question is the condition for convergence of the continued fraction $v(u)$.

18. Transforming the differential into algebraic form by the simple relation $t = -\exp(ix)$, we obtain

$$t(\gamma t^2 - t + \gamma) \frac{d^2 y}{dt^2} + (\gamma t^2 - t + \gamma) \frac{dy}{dt} + p^2 y = 0. \quad (14)$$

The equation has the singularities $t=0$, $t=\infty$, arising from the above transformation, and $t=t_q^{\pm} = (1 \pm \Delta)/2\gamma$ [where $\Delta = \sqrt{1 - 4\gamma^2}$] corresponding to the singularities of the original equation given by $\cos x = -1/2\gamma$.

If $2\gamma < 1$, the two last singularities lie on the real axis and their abscissæ are reciprocal to each other. If $2\gamma=1$, the singularities coincide at the point $t=1(x=0)$; and finally, when $2\gamma > 1$, they fall on the unit circle $|t|=1$ and are complex conjugate. (The circle $|t|=1$ obviously corresponds to the real axis in the x -plane.)

When $2\gamma < 1$, Floquet's form of the integral is transformed, in terms of t , to the form t^u (or t^{-u}) multiplied by a Laurent series converging in the annular region between the two circles $|t|=(1 \pm \Delta)/2\gamma$. Hence the t -plane in which the function $y t^{-u}$ (or $y t^u$) is defined must not be cut across this region; that is, the function $y t^{\pm u}$ (corresponding to a periodic function of x) can be uniquely defined in a t -plane, cut only between O and Q and between P and ∞ .

When $2\gamma > 1$, the width of the annular space becomes zero, and the Laurent series in t , if any (that is, the Fourier series in x), converges at most on the circle $|t|=1$, that is, on the real x -axis.

19. The algebraic form of the equation is helpful in discussing its properties in the case $2\gamma > 1$. When $\gamma=\frac{1}{2}$, the continued fraction $v(u)$ converges only if $1 - 8p^2 > 0$, that is, if $p < \frac{1}{2}\sqrt{2}$. The point $\gamma=\frac{1}{2}$, $p=\frac{1}{2}\sqrt{2}$, actually appears to be an exceptional point of the equation, which, in algebraic form, is in this case reduced to

$$t(1-t)^2 \frac{d^2 y}{dx^2} + (1-t)^2 \frac{dy}{dt} + \frac{1}{4} y = 0.$$

This well-known Legendre equation can be reduced to standard hypergeometric form by the substitution *

* When x varies on the real axis, z varies on the straight line $R(z)=\frac{1}{2}$ from $-\infty$ to ∞ .

$$s = t/(t-1) = e^{i\pi}/(1 + e^{i\pi}),$$

which gives

$$s(1-s)\frac{d^2y}{ds^2} + (1-2s)\frac{dy}{ds} - \frac{1}{4}s = 0.$$

For a complete discussion of the above equation, see Tannery and Molk, *Fonctions Elliptiques*, Paris, 1898, III, 188. We write $\mathbf{K}(s)$ for the elliptic integral having s as square of the Legendrian modulus ($s=k^2$):

$$\mathbf{K}(s) = \frac{1}{2}\pi F\left\{\frac{1}{2}, \frac{1}{2}; 1; s\right\} = \frac{1}{2}\pi\lambda(s) = \frac{1}{2}\pi\left\{1 + \left(\frac{1}{2}\right)^2s + \left(\frac{1.3}{2.4}\right)^2s^2 + \dots\right\}$$

$$\mathbf{K}'(s) = \mathbf{K}(1-s) = -\frac{1}{2}\left\{\lambda(s) \log \frac{s}{16} + 4\mu(s)\right\},$$

where $\mu(s)$ is the power-series

$$\mu(s) = \left(\frac{1}{2}\right)^2(1-\frac{1}{2})s + \left(\frac{1.3}{2.4}\right)^2(1-\frac{1}{2} + \frac{1}{2} - \frac{1}{2})s^2 + \dots$$

The equation is satisfied by the complete elliptic integrals of the first kind, $\mathbf{K}(s)$ and $\mathbf{K}'(s) = \mathbf{K}(1-s)$. In terms of t or x , two independent integrals y_1 and y_2 , being real on the circle $|t|=1$, that is, on the real x -axis, are

$$2y_1 = \sqrt{1-t}\mathbf{K}(t) + \sqrt{1-\frac{1}{t}}\mathbf{K}\left(\frac{1}{t}\right) = \sqrt{1+e^{i\pi}}\mathbf{K}(-e^{i\pi}) + \sqrt{1+e^{-i\pi}}\mathbf{K}(-e^{-i\pi}),$$

$$2iy_2 = \sqrt{1-t}\mathbf{K}(t) - \sqrt{1-\frac{1}{t}}\mathbf{K}\left(\frac{1}{t}\right) = \sqrt{1+e^{i\pi}}\mathbf{K}(-e^{i\pi}) - \sqrt{1+e^{-i\pi}}\mathbf{K}(-e^{-i\pi}).$$

The elliptic integrals are single-valued in a t -plane, cut along the segments $(\infty, 1)$ or $(1, 0)$ respectively; similarly, the radicals $\sqrt{1-t}$ and $\sqrt{1-1/t}$ have branch lines between 1 and ∞ or 1 and 0 respectively. The circle $|t|=1$ can then be described from $t=1$ to $t=1$ again, indefinitely; that is, x can vary from $-\pi$ to π , 3π , and so on, with the corresponding point $t = -e^{-i\pi}$ remaining always on the first sheet of the Riemann surface. In this case the point corresponding to $\sqrt{1+e^{i\pi}}$ repeatedly describes the right half of an ordinary lemniscate, having the points ± 1 as foci, if we choose for the square root the value with positive real part.

Correspondingly, the integrals y_1 and iy_2 , which can be defined simply as being the real and imaginary parts of $\sqrt{1+e^{i\pi}}\mathbf{K}(-e^{i\pi})$ respectively, are both periodic functions of x , with period 2π . The functions, which can easily be expressed in the form of Fourier series, are the two eigenfunctions \mathbf{K}_1 and Σ_1 , in the case $\gamma = \frac{1}{2}$. The double curve $u_0 = 1$ has then its end-point at $\gamma = \frac{1}{2}$, $p = \frac{1}{2}\sqrt{2}$.

Alternatively, we may also regard the second turn of the circle $|t|=1$ (corresponding to $\pi < x < 3\pi$) as described on the second sheet of the Riemann surface, the passage from the first to the second sheet taking place at the singular point $t=1$ ($x=\pm\pi$), which is the end-point of the cuts.

The point corresponding to $\sqrt{1+e^{is}}$ then describes the left half of the lemniscate, while $K(-e^{is})$ assumes the conjugate value $K(-e^{-is})$. This corresponds to the assumption of a pair of integrals defined as follows:—

for

$$-\pi < x < \pi: \quad R\{\sqrt{1+e^{is}}K(-e^{is})\} = y_1 \quad \text{and} \quad -iR\{\sqrt{1+e^{is}}K(-e^{is})\} = y_2,$$

for

$$\pi < x < 3\pi: \quad R\{-\sqrt{1+e^{is}}K(-e^{is})\} = -y_1 \quad \text{and} \quad -iR\{-\sqrt{1+e^{is}}K(-e^{is})\} = y_2,$$

the same branch of the square root being chosen in all cases. The integral of the differential equation defined as y_1 for the first half of the period, and as $-y_1$ for the second half, is actually a "half-periodic" function. It is the second eigenfunction K_1 of § 14, and can easily be expanded as a Fourier series, which in point of fact proves to be an odd cosine series. The integral (y_2, y_2) , on the contrary, is a mere duplication of the integral Σ_1 .

We thus conclude that the end-point of the curve-locus of the odd eigenfunction $K_1(x)$ is $\gamma=\frac{1}{2}$, $\rho=\frac{1}{2}\sqrt{2}$, which is also the terminal point of the double curve of eigenvalue $u_0=1$.

20. The general case $\rho \neq \frac{1}{2}\sqrt{2}$ can be treated in a similar way. If we write δ for $\sqrt{1-8\rho^2}$, the transformation $t = -e^{is}$ changes the equation

$$(1 + \cos x) \frac{d^2 y}{dx^2} + \rho^2 y = 0$$

into the equation in algebraic form

$$t(1-t)^2 \frac{d^2 y}{dt^2} + (1-t)^2 \frac{dy}{dt} + 2\rho^2 y = 0,$$

and this is the Riemann P -equation

$$P \left\{ \begin{array}{ccc} 0 & \infty & 1 \\ 0 & 0 & \frac{1+\delta}{2} \\ 0 & 0 & \frac{1-\delta}{2} \end{array} \quad t \right\}.$$

It is reduced to the hypergeometric form by the usual transformation

$$s = t/(t-1), \quad t = s/(s-1),$$

giving

$$s(1-s)\frac{d^2y}{ds^2} + (1-2s)\frac{dy}{ds} - 2p^2y = 0.$$

The equation is symmetric in s and $1-s$, as in the particular case $2p^2 = \frac{1}{4}$. It is therefore convenient to assume, as independent integrals, the following forms, which reduce to $\mathbf{K}(s)$ and $\mathbf{K}'(s) = \mathbf{K}(1-s)$ when $\delta = 0$, and still satisfy the fundamental equation $\mathbf{k}'(s) = \mathbf{k}(1-s)$:—

$$\mathbf{k}(s) = \frac{1}{2}\pi \sec \frac{1}{2}\pi\delta \cdot F\left\{\frac{1}{2}(1+\delta), \frac{1}{2}(1-\delta); 1; s\right\},$$

where

$$\begin{aligned} F\left\{\frac{1}{2}(1+\delta), \frac{1}{2}(1-\delta); 1; s\right\} &= \bar{\lambda}(s) = 1 + \lambda_1 s + \lambda_2 s^2 + \dots \\ &= 1 + \frac{1-\delta^2}{2^2}s + \frac{(1-\delta^2)(3^2-\delta^2)s^2}{(2\cdot 4)^2} + \dots \end{aligned}$$

and

$$\mathbf{k}'(s) = -\frac{1}{2}\{\bar{\lambda}(s)[\log s - 4C] + 4\bar{\mu}(s)\},$$

$\bar{\mu}(s)$ here denoting the series

$$\bar{\mu}(s) = \lambda_1\mu_1 s + \lambda_2\mu_2 s^2 + \lambda_3\mu_3 s^3 + \dots$$

where

$$\mu_1 = \frac{1}{1-\delta^2} - \frac{1}{2}; \quad \mu_2 = -\frac{1}{1-\delta^2} - \frac{1}{2} + \frac{3}{3^2-\delta^2} - \frac{1}{4}; \quad \dots$$

$$C = \mu_{\infty} = -\frac{1}{2}\gamma + \frac{1}{2}\pi \tan \frac{1}{2}\pi\delta - \frac{1}{2}\Psi\left\{\frac{1}{2}(1+\delta)\right\},$$

γ here denoting Euler's constant, and $\Psi(x) = \Gamma'(x)/\Gamma(x)$. When $\delta = 0$, C reduces to $\log 2$, so that $\mathbf{k}'(s)$ reduces to $\mathbf{K}'(s)$, as it should.

Two real independent integrals, valid on the circle $|t| = 1$ (the singular point $t = 1$ excepted), that is, on the real x -axis (the points $x = (2n+1)\pi$ excepted), or on the line $R(s) = \frac{1}{2}$ (the point at infinity excepted), are, as above,

$$2iy_1 = \mathbf{k}(s) + \mathbf{k}'(s), \quad 2iy_2 = \mathbf{k}(s) - \mathbf{k}'(s),$$

since $\mathbf{k}(s)$ and $\mathbf{k}'(s) = \mathbf{k}(1-s)$ are conjugate to each other when s is on the line $R(s) = \frac{1}{2}$. The functions $\mathbf{k}(s)$ and $\mathbf{k}'(s)$ can be written as definite integrals, using Euler's expression for the hypergeometric series, as follows:—

$$\mathbf{k} = \int_0^{\frac{1}{2}\pi} \left[\frac{\cos \tau}{\sin \tau \operatorname{dn} \tau} \right]^2 d\tau; \quad \mathbf{k}' = -ie^{\frac{1}{2}\pi i} \int_0^{\frac{1}{2}\pi} \left[\frac{\cos \tau}{\sin \tau \operatorname{dn} \tau} \right]^2 d\tau.$$

The integrals can be regarded as (transcendent) modular functions, which remain unchanged for all those transformations of the ratio of the periods that leave the Legendrian modulus unchanged. They are thus apparently periodic functions of x , since the increase of x by 2π simply increases the ratio of the periods by 2.

Remark.—Before coming to any conclusion concerning the existence of periodic solutions for any value of p , when $2\gamma > 1$, one would have to investigate more carefully the behaviour of the integrand, in particular its dependence on the index δ , which may assume any value, real or complex. The existence of periodic solutions for any value of p is put in better evidence by the discussion of the case $2\gamma > 1$ in § 21; the procedure set out there can also be applied to the present case.

It is concluded that the first eigencurve $u_0 = \frac{1}{2}$ ends at $\gamma = \frac{1}{2}$, $p = 0$; the second eigencurve $u_0 = \frac{1}{2}$, and the double curve $u_0 = 1$, at $\gamma = \frac{1}{2}$, $p = \frac{1}{2}\sqrt{2}$. It is regarded as very probable that *all* the eigencurves may terminate at this particular point, since the resonance equation is divergent for greater values of p . If this should be the case, the half-straight-line $\gamma = \frac{1}{2}$, $p = \frac{1}{2}\sqrt{2}$ might be regarded as the limiting form of the eigencurve $u_0 = n$ as n tends to infinity, thus permitting the existence of periodic solutions at all its points.

IX. THE CASE $2\gamma > 1$

21. When $2\gamma > 1$ equation (2) is not a Hill's equation, so that the determination of the characteristic exponent u_0 is possible neither by Hill's method of the determinantal equation, nor by the method of continued fractions as used in the earlier sections of this paper.

Since the singularities of the equation, expressed in its algebraic form, all lie on the unit circle, it is known that the periodic solutions, if any, or the expressions in Floquet's form $e^{i u_0 x} F(x)$, where $F(x)$ is a Fourier series, are valid only on the real x -axis.

The determination of u_0 is possible, however, according to Floquet's theory, as soon as we are able to express the value of two integrals for $x = x_0 + 2\pi$ in terms of their values at $x = x_0$.

The equation, reduced to algebraic form, is still (14). The two proper singularities, $t = P = (1 + \Delta)/2\gamma$ and $t = Q = (1 - \Delta)/2\gamma$, are on the unit circle $|t| = 1$, where they are represented by the conjugate points $P = e^{i\lambda}$ and $Q = e^{-i\lambda}$, where $2\gamma \cos \lambda = 1$.

Since the equation is *reciprocal* in the sense that the substitution $1/t$ for t leaves it unchanged, the existence of an integral $f(t)$ defined around $t = 0$ implies the existence of an integral $f(1/t)$ valid around $t = \infty$. The characteristic exponents at the four singularities are all integers, namely

0, 0 at $t=0$ and $t=\infty$, and 0, 1 at $t=P$ and $t=Q$, and the irregular integrals there are all logarithmic (except at $t=P$ and $t=Q$ when $p=0$).

In the neighbourhood of $t=0$ we immediately construct the two independent integrals:

$$g_0(t) = 1 + g_1 t + g_2 t^2 + g_3 t^3 + \dots$$

$$h_0(t) = g_0 \cdot \log t + h_1 t + h_2 t^2 + h_3 t^3 + \dots$$

where the constants g_n and h_n are given by the recurrence relations

$$\gamma(n+1)g_{n+1} - (n^2 - p^2)g_n + \gamma(n-1)g_{n-1} = 0, \quad g_0 = 1, g_{-1} = 0,$$

$$\gamma(n+1)h_{n+1} - (n^2 - p^2)h_n + \gamma(n-1)h_{n-1} = -2\{\gamma(n+1)g_{n+1} - ng_n + \gamma(n-1)g_{n-1}\}.$$

The series converge inside the circle $|t| = 1$; the integrals $g_\infty(t) = g_0(1/t)$ and $h_\infty(t) = h_0(1/t)$ converge for $|t| > 1$ and represent two independent solutions valid around $t=\infty$. The integrals h are single-valued in the t -plane, cut, for example, from $-\infty$ to 0.

Around the singularity $t=P=(1+\Delta)/2\gamma$, we similarly define a regular integral

$$g_P = p_1(t-P) + p_2(t-P)^2 + \dots,$$

where the constants p_n are defined by

$$(n+1)nP\Delta p_{n+1} + \{n(n-1)\gamma P + n^2\Delta + p^2\}p_n + (n-1)^2\gamma p_{n-1} = 0$$

and it is convenient to assume $p_1 = Q = e^{i\lambda}$. Similarly around $t=Q$ we have the regular integral *

$$g_Q = q_1(t-Q) + q_2(t-Q)^2 + \dots$$

with constants q_n defined by

$$-(n+1)nQ\Delta q_{n+1} + \{n(n-1)\gamma Q - n^2\Delta + p^2\}q_n + (n-1)^2\gamma q_{n-1} = 0,$$

and we agreed to assume $q_1 = P = e^{i\lambda}$. Since P and Q , as well as the coefficients p_n and q_n , are conjugate, we easily see that

$$g_P(t) = \bar{g}_Q(t),$$

where \bar{x} denotes the complex conjugate of x .

Around $t=P$ and similarly around $t=Q$ we can also define the irregular integrals

$$h_P = g_P \cdot \log(t-P) + r_0 + r_1(t-P) + r_2(t-P)^2 + \dots,$$

* The integral g_Q is obtained from g_P by substituting $-\Delta$ for Δ . Since Δ is purely imaginary, the coefficients of g_Q are conjugate to those of g_P .

with a similar expression for h_Q , but these are of much more restricted interest in the present context. We may merely remark that, although these integrals are singular at $t=P$ and $t=Q$ respectively, they are finite in magnitude at these points.*

The expressions $g_P(1/t)$ and $g_Q(1/t)$, convergent about $t=Q$ and $t=P$ respectively, are also integrals of the equation; and it is easily seen that, with the values assumed for the arbitrary coefficients p_1 and q_1 , we have the simple relation

$$g_P(1/t) = -g_Q(t).$$

Hence, when t is on the unit circle, so that $\bar{t} = 1/t$, g_P and g_Q are purely imaginary.

The integrals g_P and g_Q , which have regions of convergence in common with the integrals g_0 and h_0 , can be expressed as linear combinations of the latter with constant (complex) coefficients. It is seen at once that if $g_P = ag_0 + bh_0$ we shall have $g_Q = \bar{a}g_0 + \bar{b}h_0$.

The complex constants a and b can be determined, when required, by computing the (complex) series g_P and the integrals g_0 and h_0 , as well as their derivatives, at a point where all the expressions converge; for instance, at the point $t = 1/2\gamma$, which certainly belongs to the regions of convergence of all the series.

22. If we suppose that the above linear relations are known, we can then determine the variation of the integrals g_P and g_Q for a 2π -increase in the original variable x ; that is, for a positive turn round the origin of the actual variable t . When t describes a closed circuit round the origin $g_0(t)$ remains unchanged, and $h_0(t)$ increases by $2\pi ig_0$. By eliminating g_0 and h_0 between the expressions for g_P and g_Q and those of the new values g_P^* and g_Q^* , we easily obtain †

$$g_P^* = a_{11}g_P + a_{12}g_Q = \left\{ 1 + 2\pi i \frac{b\bar{b}}{ab - db} \right\} g_P - 2\pi i \frac{b^2}{ab - db} g_Q,$$

$$g_Q^* = a_{21}g_P + a_{22}g_Q = 2\pi i \frac{b^2}{ab - db} g_P + \left\{ 1 - 2\pi i \frac{b\bar{b}}{ab - db} \right\} g_Q.$$

A linear combination $F(x) = Ag_P(x) + Bg_Q(x)$ (where g_P and g_Q are regarded as functions of x by means of the relation $t = e^{-ix}$) will then satisfy Floquet's condition

$$F(x + 2\pi) = \rho F(x),$$

* We have $h_P(P) = -h_Q(Q) = -\Delta/p^2$.

† The determinant of the transformation (which is the ratio of the two Wronskian determinants of the functions g_P^* , g_Q^* and g_P , g_Q respectively) is equal to 1, by a theorem due to Poincaré.

provided that $Ag_P^* + Bg_Q^* = \rho(Ag_P + Bg_Q)$. From the expressions just given for g_P^* and g_Q^* , we find, using $a_{11}a_{22} - a_{12}a_{21} = 1$, that the only possible values of ρ are those given by the quadratic equation

$$\rho^2 - (a_{11} + a_{22})\rho + 1 = 0.$$

Since in the present case we *always* have $a_{11} + a_{22} = 2$, so that the equation has the double root $\rho = 1$, we conclude that for *any value of the constants p and γ of the original equation* (provided that $2\gamma > 1$) it is always possible to obtain two independent integrals, *both periodic with period 2π* .

This conclusion is generally valid whenever (i) the singularities, other than 0 and ∞ , all lie on the circle $|t| = 1$; (ii) the singularity is logarithmic. The conclusion is also true for other values of the characteristic exponents at $t=0$, subject to conditions which are not difficult to formulate.

23. Two values of the constants A and B making $\rho = 1$ (that is, such that u_0 is an integer) are $A = b$, $B = -b$. For the corresponding Floquet solution we obtain

$$F(x) = bg_P - bg_Q = Cg_0.$$

This means that g_0 , when regarded as a function of x , can be expressed as a Fourier series valid on the real x -axis, corresponding to the circle $|t| = 1$. Since g_0 is also defined, and single-valued, *inside* the circle $|t| = 1$, the corresponding Fourier series will be valid in the upper half-plane $I(x) > 0$.

The second periodic integral will obviously be $g_0(1/t) = g_\infty(t)$, defined for $|t| > 1$, and representing a Fourier series in x , valid in the lower half of the x -plane. Since g_0 and g_∞ are conjugate to each other when $|t| = 1$, two real integrals valid for $|t| = 1$ are obtained by considering the real and the imaginary parts of $g_0(t)$ respectively. They give, in terms of x , two real Fourier series, which are valid only on the real x -axis.

The validity of the series $g_0(t)$ for $|t| = 1$ can also be proved directly. Since its coefficients satisfy the relation

$$\gamma(n+1)^2 g_{n+1} - (n^2 - p^2)g_n + \gamma(n-1)^2 g_{n-1} = 0,$$

the expression $V(n) = n^2 g_n$, considered as a function of n , satisfies the usual difference equation

$$\gamma V(n+1) - \left(1 - \frac{p^2}{n^2}\right)V_n + \gamma V(n-1) = 0,$$

which tends asymptotically to

$$\gamma V(n+1) - V(n) + \gamma V(n-1) = 0.$$

Two independent (complex) integrals of the above equation are $P^n = e^{in\lambda}$ and $Q^n = e^{-in\lambda}$. Two independent integrals in real form are consequently $\cos n\lambda$ and $\sin n\lambda$. Since g_n is always real, $n^2 g_n$ will tend to a linear combination of $\cos n\lambda$ and $\sin n\lambda$. Such expressions have no definite limit (that is, the expression $n^2 g_n$ does not converge) but are limited in magnitude. This means that $|g_n| < K/n^2$, K being a suitable constant.

It follows that the series g_n tends asymptotically to a series the terms of which are respectively smaller in modulus than those of the series $K \sum 1/n^2$, which converges absolutely for $|\lambda| = 1$. The convergence of the Fourier series is actually slow, so that a method giving a closed expression for its sum is much to be desired.

X. GENERAL CONCLUSIONS AND SUMMARY

24. When $2\gamma < 1$, the equation considered, that is, the equation

$$\frac{d^2 y}{dx^2} + \frac{p^2}{1 + 2\gamma \cos x} y = 0,$$

has periodic solutions, with period 2π or 4π , only in those cases where the parameters p and γ are such that the representative point belongs to a curve where $2u_0$ is an integer. In the regions of stability bounded by such curves, with $2u_0$ an odd integer, an infinity of curves exist giving periodic solutions with periods that are multiples of 2π .

Finally, regions of lability also exist, where the equation has no stable solutions. With the exception of the first one, bounded by the two curves $u_0 = \frac{1}{2}$, starting at $p = \frac{1}{2}$, $\gamma = 0$, the regions of lability are extremely narrow; the region comprised between the curves $u_0 = 3/2$ is already of negligible width.

When periodic solutions exist, they are given by Fourier series valid in a strip of the x -plane, the width of which increases with decreasing γ .

In the cases $2\gamma > 1$, on the contrary, stable solutions exist everywhere; they are periodic with period 2π , and are represented by Fourier series converging in a strip of zero width, that is, on the real axis only. All the solutions are finite in magnitude at the singular, real points given by $2\gamma \cos x + 1 = 0$, but in general their first derivative has a logarithmic infinity there, so that the Fourier series cannot be differentiated term by term.

If the equation is written in the general form

$$(a + b \cos x) \frac{d^2 y}{dx^2} + y = 0, \quad a = 1/p^2, \quad b = 2\gamma/p^2,$$

the positive half-quadrant $b > a$ is a region of stability; in the lower half of the quadrant the regions of stability are bounded by curves which

are tangent to the straight line $a=b$ at the point $a=b=8$, with the exception of one of the curves $u_0=\frac{1}{2}$, which has the line $a=b$ as asymptote.

The argument of § 20 does not establish this statement finally, though it may be regarded as very probable. The double curve $u_0=1$, and one of the curves $u_0=\frac{1}{2}$, actually converge to this point.

For the general treatment of linear equations with periodic coefficients, and the questions of stability and lability, we may cite the works given in the list below.

REFERENCES TO LITERATURE

- NÖRLUND, N. E., 1924. *Vorlesungen über Differenzenrechnung*, Berlin.
SANSONE, G., 1941. *Le equazioni differenziali nel campo reale*, Bologna. Esp. Chapter VI.
STRUTT, M. J. O., 1932. *Lamé-sche, Mathieu-sche, und verwandte Funktionen*, Berlin. This contains an extensive bibliography on Hill's problem.
——, 1943. *Eigenwaarde-Krommen bij Problemen van Hill*, I and II, Amsterdam.
——, 1943. *Eigenfuncties bij Problemen van Hill*, I and II, Amsterdam.
TANNERY, J., and MOLK, J., 1898. *Fonctions Elliptiques*, Paris.
WHITTAKER, E. T., and WATSON, G. N., 1927. *Modern Analysis*, Cambridge. Esp. Chapter XIX.

(Issued separately June 16, 1950)

IV.—Studies in Practical Mathematics. V. On the Iterative Solution of a System of Linear Equations.* By **A. C. Aitken**, D.Sc., F.R.S., Mathematical Institute, 16 Chambers Street, Edinburgh, 1.

(MS. received April 28, 1949. Read July 4, 1949)

SYNOPSIS

The convergence of customary processes of iteration for solving linear equations, in particular simple and Seidelian iteration, is studied from the standpoint of matrices. A new variant of Seidelian iteration is introduced. In the positive definite case it always converges, the characteristic roots of its operator being real and positive and less than unity.

1. PRELIMINARY CONSIDERATIONS

THERE has appeared recently a series of papers (Bodewig, 1947) reviewing various extant methods of solving simultaneous linear equations and assessing their relative efficiency. It is claimed, on the basis of this assessment, that the oldest and most elementary of these methods, that of successive elimination of the unknowns followed by resubstitution, involves fewer and simpler operations than any of its more recent competitors. The assessment of efficiency is based on the number of operations required before the solution is complete; for example, it is stated that in the case of a system of n equations, the whole process of solution by the method of elimination requires $n(n^2 + 3n - 1)/3$ multiplications and $n(n - 1)(2n + 5)/6$ additions.

These assessments must be viewed with respect; but are equally subject to qualification and revision. For example, many operations of addition are not separate from those of multiplication, but on the machine are cumulated along with them. Again, copying down is itself an operation that takes a relatively large proportion of time and involves a risk of error; any operation that minimizes the necessity of such copying is advantageous. There are certainly cases, especially when the matrix A of the system is dominated by its diagonal elements, in which Seidelian iteration converges well. Further, if in such a case the greatest characteristic root of the iterative operation is real and rather small compared with 1, powerful methods are available for gaining at one step a much enhanced approximation. Finally, the Southwellian

* This paper was assisted by a grant from the Carnegie Trust for the Universities of Scotland.

technique, which is that of Seidelian iteration with an admixture of opportunism, has stood the test of trial in the engineering applications with which it deals.

In the present paper, we first make some general observations on the rapidity of convergence of iterative processes (some of this crosses ground already traversed by Dr Bodewig), and then go on to introduce a new variant of the Seidelian process, based on an operator which, when the matrix A of the system to be solved is positive definite, has characteristic roots real and confined to the range $0 < \lambda < 1$. The advantages of this will be seen in due course.

For the sake of economy of space, the numerical examples refer to systems of low order, and are indeed capable of being solved just as rapidly in other ways. They are intended to serve as illustrations of principle, and are designed to exhibit to the eye, in a not specially favourable case, those properties of convergence that are derived in the text.

2. THE GENERAL PROCESS OF ITERATION

The usual processes of iteration have often been described (*e.g.* in Frazer, Duncan and Collar, 1938; Bodewig, 1947) and are easily classified. Let the system of equations be denoted in matrix notation by $Ax = h$. The matrix A can be expressed in infinitely many ways as $B - C$. We choose B as a non-singular matrix such that either B^{-1} is easily evaluated as a first step, or else the effect of B^{-1} on any vector can be obtained by simple arithmetical routine. The choice of B is wide; in the simplest case it could be scalar; in the next simplest, diagonal; in the next simplest, upper or lower triangular. Iteration consists in the use of the recurrence relation

$$Bx^{(t+1)} = Cx^{(t)} + h, \quad \text{that is,} \quad x^{(t+1)} = B^{-1}Cx^{(t)} + B^{-1}h, \quad (1)$$

the best working rule consisting in the use of vector-differences, thus,

$$B\{x^{(t+1)} - x^{(t)}\} = C\{x^{(t)} - x^{(t-1)}\}. \quad (2)$$

The iterated vectors $x^{(t)}$ are derived in this way from an initial vector $x^{(0)}$, adopted as a first approximation to the vector x of solutions.

The rapidity of convergence of the sequence $x^{(t)}$ towards x thus depends on the latent roots of the matrix $B^{-1}C$. The characteristic equation of the iteration is therefore $|\lambda B - C| = 0$, and if its roots, in descending order of moduli, are $\lambda_1, \lambda_2, \dots, \lambda_n$, convergence will be assured provided that $|\lambda_1| < 1$, and will be the more rapid the smaller $|\lambda_1|$ is. In certain cases λ_1 will be real; if this can be ensured, methods for accelerating the

convergence can be applied. These methods can, in fact, be applied even when $|\lambda_1| > 1$, provided that $|\lambda_2| < 1$, in much the same way that $1 + \lambda + \lambda^2 + \dots + \lambda^{t-1}$, when provided with the remainder term $\lambda^t/(1-\lambda)$, will yield $1/(1-\lambda)$ for *all* values of λ ; but this case has little practical value. If λ_1 is one of a pair of conjugate complex roots, accelerative methods exist (Aitken, 1925, p. 302), but are less convenient to apply.

A first classification of iterative methods may be based on the nature of B . If B is purely diagonal, for example if it is the "vertebra" of A , so that $b_{ii} = a_{ii}$, we have the type of iteration that is often called *simple* or *ordinary*, but could equally well be called *diagonal*. If B is "lower triangular", that is, if $b_{ij} = 0$, $i < j$, $b_{ii} \neq 0$, we have what may be called *lower triangular* iteration. Not essentially different is *upper triangular* iteration. Seidelian iteration (Seidel, 1874), as usually understood, is the special lower triangular iteration where $b_{ii} = a_{ii}$, $b_{ij} = a_{ij}$, $i > j$, $b_{ii} = 0$, $i < j$; or any similar iteration with the equations in some permuted order. A variant due to Morris (Frazer, Duncan and Collar, 1938, p. 132) uses a preliminary evaluation of B^{-1} and $B^{-1}h$, but the results at each step are the Seidelian ones.

At this stage we may illustrate, by simple examples of the 3rd order, the characteristic equations of the simplest diagonal iteration ($b_{ii} = a_{ii}$) and of lower triangular Seidelian iteration respectively:

$$\begin{vmatrix} \lambda a_{11} & a_{12} & a_{13} \\ a_{21} & \lambda a_{22} & a_{23} \\ a_{31} & a_{32} & \lambda a_{33} \end{vmatrix} = 0, \quad (3)$$

$$\begin{vmatrix} \lambda a_{11} & a_{12} & a_{13} \\ \lambda a_{21} & \lambda a_{22} & a_{23} \\ \lambda a_{31} & \lambda a_{32} & \lambda a_{33} \end{vmatrix} = 0. \quad (4)$$

If A is symmetric, the roots of (3) are real, though not necessarily all less than 1; in (4), on the other hand, where we note the unsymmetrical disposition of the λ 's, the roots may be real or complex according to circumstance, even when A is symmetric. In justification of this last remark we may refer to the simple positive definite example of the 3rd order, $a_{ii} = 3$, $a_{ij} = a_{ji} = 2$, $i \neq j$.

In simple iteration it is known (Bodewig, 1947), and will be proved here, that if A is symmetric and *positive definite*, it is always possible to find a diagonal matrix B such that the numerically greatest latent root λ_1 of the iterating operator is not only real but such that $|\lambda_1| < 1$. It is also known (Whittaker and Robinson, 1929, p. 255; Bodewig, 1947) that in this positive definite case Seidelian iteration always converges. We proceed to prove these facts from the standpoint of matrix theory.

3. THE POSITIVE DEFINITE CASE

Let A be positive definite. Then its diagonal elements a_{ii} are all positive and constitute a diagonal matrix D . We may normalize the given equations $Ax = k$ to

$$D^{-1}AD^{-1}y = D^{-1}k, \text{ where } y = D^{\frac{1}{2}}x. \quad (1)$$

This normalization is for the purpose of theory only, and need not be resorted to in practice. There, it is enough to semi-normalize, obtaining the non-symmetric system $D^{-1}Ax = D^{-1}k$. However, since

$$D^{-1}A = D^{-\frac{1}{2}}(D^{-\frac{1}{2}}AD^{-\frac{1}{2}})D^{\frac{1}{2}}, \quad (2)$$

the matrices $D^{-1}A$ and $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ are similar and so have the same latent roots. Each has unit diagonal elements, and so the trace is equal to n ; hence, the latent roots being all positive, we must have $\lambda_1 < n$, $\lambda_n > 0$. By writing $D^{-1}A = \frac{1}{2}n(I - C)$, $k = \frac{1}{2}nk$, we reduce the equations to the form $(I - C)x = k$, where the latent roots of C evidently lie in the range $-1 < \lambda < 1$. A simple iteration based upon

$$x^{(t+1)} = Cx^{(t)} + k \quad (3)$$

can now be applied. The process will always converge; though often, as we shall see (§ 6) by an example, with disappointing slowness. Naturally, if we knew beforehand the approximate location of the latent roots of $D^{-1}A$, and especially of the smallest root, we could in most cases make a better change of matrix origin than $\frac{1}{2}nI$; but such knowledge is not usually precise enough.

Let us next consider the Seidelian case. The property of convergence, in the case when A is positive definite, is perhaps most easily deduced indirectly, from the fact that $Ax = k$ can be regarded, and in infinitely many ways, as the normal equations of some linear least-square problem. For we have $A = M'M$, $|M| \neq 0$, a resolution which is possible in infinitely many ways, since HM , where H is an arbitrary orthogonal matrix, can replace M here; and then the equations $M'Mx = k$ are the normal equations corresponding to the "observational" equations $Mx = (M')^{-1}k$.

These normal equations arise from the minimizing of the definite quadratic form $s^2 = (Mx - k)'(Mx - k)$, where $k = (M')^{-1}k$, and this is a sum of squares which can be transformed to a different sum of squares (Whittaker and Robinson, 1929, p. 255) by the classical reduction of Lagrange (Turnbull and Aitken, 1932, p. 83), in which, if we gather all terms in x_1 into a squared term, we obtain

$$s^2 = a_{11}^{-1}(a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n - k_1)^2 + \dots, \quad (4)$$

every term later than the first being free of x_1 . To annul the squared residual bracketed in the first term is to reduce s^2 ; and such annulling is the typical single operation in any phase of Seidelian iteration. The point is, that the annullment is effected by modifying x_1 only, and so the later terms in s^2 are unaffected. Each operation of Seidelian iteration thus reduces s^2 ; and so the vector of values $\{x_1 x_2 \dots x_n\}^{(t)}$ converges, as desired, to those values that minimize s^2 . It follows, *a posteriori*, that the latent roots λ_i of the Seidelian operation must be such that $|\lambda_i| < 1$; but, as has been seen from the simple example of § 2, they may be complex.

The Seidelian operation can be described thus. Take $d_{ii} = a_{ii}$ as before. Let C be the lower triangular matrix $c_{ii} = 0$, $c_{ij} = a_{ij}$, $i > j$. Then $A = D - C - C'$, the iterating matrix is $(D - C)^{-1}C'$, and the characteristic equation of the iteration is $|\lambda(D - C) - C'| = 0$. We learn therefore that if $D - C - C'$ is positive definite, then the roots of the above equation are such that $|\lambda| < 1$.

4. THE ACCELERATION OF CONVERGENCE

The successive vector increments are derived, as we have seen in § 2, by the recurrence

$$x^{(t+1)} - x^{(t)} = B^{-1}C\{x^{(t)} - x^{(t-1)}\}. \quad (1)$$

We are therefore on the familiar and well-explored ground of the repeated matrix operation on a vector, and we know that in the case that is being considered

$$x^{(t+1)} - x^{(t)} = \lambda_1 \{x^{(t)} - x^{(t-1)}\} \quad (2)$$

will hold with greater and greater approximation, and the more if $|\lambda_2|/|\lambda_1|$ is small. In practice it is advantageous to form the vector-difference at the earliest possible stage and to use it as operand, thus dismissing h from consideration. We shall therefore use

$$B\Delta x^{(t)} = C\Delta x^{(t-1)}, \text{ where } \Delta x^{(t)} = x^{(t+1)} - x^{(t)}, \quad (3)$$

as the recurrence. When the iterated differences are adequately small, we cumulate them upon $x^{(0)}$; and the errors of rounding-off likely to be incurred in such a cumulation can be obviated by retaining one or two additional digits, the customary expedient for "stabilizing" any calculation.

In this positive definite case, however, provided that λ_1 is real, it will usually not be necessary to continue the iteration until the vector-differences are as small as desired; it will be enough that the corresponding

elements in consecutive vector-differences should begin to show an approach to a geometrical progression, which will be of common ratio λ_1 . The accelerative methods are then available. Suppose in fact that $\Delta x_i^{(t-1)}$ and $\Delta x_i^{(t)}$ are two such corresponding elements; write $\Delta^2 x_i^{(t-1)} = \Delta x_i^{(t)} - \Delta x_i^{(t-1)}$, and form the quotient

$$\{\Delta x_i^{(t)}\}^2 / \Delta^2 x^{(t-1)}. \quad (4)$$

This (Aitken, 1925, p. 301; Holme, 1932; Steffensen, 1933) will approximate to the remainder term; it should be applied, therefore, to each x_i in the vector of solutions derived by cumulating up to $x^{(t)}$, and the resulting vector can then be tested by further Seidelian iteration. A somewhat less accurate remainder term is given by the quotient

$$\{\Delta x_i^{(t-1)} \Delta x_i^{(t)}\} / \Delta^2 x^{(t-2)}, \quad (5)$$

and this will serve sometimes as a check.

5. A MODIFICATION OF SEIDELIAN ITERATION

It has been established that in the symmetric and positive definite case Seidelian iteration converges. If a type of Seidelian iteration could be devised such that all the roots of the characteristic equation were *real*, the accelerative methods would in every case be available, and would enhance the already existing advantage of convergence. We suggest therefore the following procedure.

As before, we have $A = D - C - C'$, positive definite. Let us begin Seidelian iteration as usual, obtaining from an initial vector $x^{(0)}$ the improved values of the unknowns, $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$. Then, instead of beginning (as is the usual procedure) a new cycle $x_1^{(2)}, x_2^{(2)}, \dots$, let us go *back through the unknowns in reverse order*, $x_{n-1}^{(2)}, x_{n-2}^{(2)}, \dots, x_1^{(2)} = x_1^{(3)}$; then down again, $x_n^{(3)}, x_{n-1}^{(3)}, \dots, x_n^{(3)} = x_n^{(4)}$; then up again, and so on. This is a to-and-fro or two-phase Seidelian iteration, and the results are obtained in the order indicated below: *e.g.*

$$\begin{array}{c|cccccc} & x_1^{(1)} & x_1^{(2)} & = & x_1^{(3)} & x_1^{(4)} & = & x_1^{(5)} \\ x_2^{(0)} & \downarrow & x_2^{(1)} & \uparrow & x_2^{(2)} & \downarrow & x_2^{(3)} & \uparrow & x_2^{(4)} & \downarrow \\ x_3^{(0)} & & x_3^{(1)} & \uparrow & x_3^{(2)} & \downarrow & x_3^{(3)} & \uparrow & x_3^{(4)} & \downarrow \\ x_4^{(0)} & & x_4^{(1)} & = & x_4^{(2)} & & x_4^{(3)} & = & x_4^{(4)} & \end{array} \quad (1)$$

The two stages, down and up, are to be regarded as the complementary halves of a double Seidelian operation, rather like the two half-oscillations of a complete oscillation; so that it will be useful to speak of elements

$x_i^{(t)}$, $x_i^{(t+2)}$, $x_i^{(t+4)}$, . . . as being "in phase". Forming therefore the differences of elements in phase, let us say

$$\Delta x_i^{(t)} = x_i^{(t+2)} - x_i^{(t)}, \quad (2)$$

we can regard the corresponding vector-differences as operands, and the complete operation is then characterized by the matrix

$$(D - C')^{-1}C.(D - C)^{-1}C'. \quad (3)$$

It will now be shown that every latent root of this matrix is real and such that $0 < \lambda < 1$. For in the first place $|\lambda| < 1$, since each individual operation is of Seidelian type. Next, we have

$$(D - C')^{-1}C.(D - C)^{-1}C' = D^{-1}\{(I - K')^{-1}K(I - K)^{-1}K'\}D^1, \quad (4)$$

where $K = D^{-1}CD^{-1}$. This again, since K and $(I - K)^{-1}$ are permutable, is similar to $(I - K)^{-1}KK'(I - K')^{-1}$, namely a non-negative definite matrix of form $M'M$ and in general of rank $n - 1$, since K is in general of rank $n - 1$. We conclude that $0 < \lambda < 1$. The effective positiveness of the roots confers some arithmetical advantage.

It may be of interest to derive the above result from first principles. To each of the latent roots λ there corresponds a non-trivial characteristic vector q such that

$$(I - K')^{-1}(I - K)^{-1}KK'q = \lambda q. \quad (5)$$

Hence

$$q'KK'q = \lambda q'(I - K)(I - K')q,$$

that is,

$$\lambda = q'KK'q / \{q'(I - K - K')q + q'KK'q\}. \quad (6)$$

The quadratic form $q'(I - K - K')q$ is positive definite, by hypothesis, and $q'KK'q$ is non-negative definite. It follows that $0 < \lambda < 1$.

In the practical technique we begin with $x^{(0)}$, form $x^{(1)}$ with the downward operations and then $x^{(2)}$ with the upward ones. Constructing then the vector-difference $x^{(2)} - x^{(0)}$, we take it as operand for the further downward and upward operations. When convergence seems adequate, we can cumulate in two ways, first upon $x^{(0)}$ with all vector differences that are in phase with it, and again upon $x^{(1)}$ with all vector-differences in phase with it; these two vectors of iterated solutions can be used to check each other.

6. NUMERICAL EXAMPLES OF THE VARIOUS ITERATIONS

To solve

$$\begin{bmatrix} 3.17 & 0.92 & -1.07 & 1.13 \\ 0.92 & 3.86 & -0.89 & -0.77 \\ -1.07 & -0.89 & 5.14 & 1.79 \\ 1.13 & -0.77 & 1.79 & 6.23 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 8.08 \\ 6.32 \\ 5.58 \\ 11.05 \end{bmatrix},$$

beginning with values 1.5, 1.5, 1.0 as first approximations for x_2, x_3, x_4 . The usual Seidelian iteration gives the first iterated vector {2.2634 1.6432 1.4931 1.1372}. Forming the first vector-difference, we continue:

							Check
	-928	-479	-151	-52	-27	2.0997	8.0796
1432	479	58	15	4	1	1.6989	6.3200
- 69	-588	-228	- 85	-29	-15	1.3986	5.5797
1372	396	160	54	18	9	1.2009	11.0496

Stopping at the stage where each entry seems to be about one-third of the corresponding entry in the column before, we apply the corrections of § 4 (4), for example $-52/99 = -27$, and cumulate upon the earliest values 2.2634, 1.5, 1.5 and 1.0. The check suggests that the solutions have at most small errors in the 4th decimal place. The accurate values are in fact {2.0999 1.6989 1.3987 1.2009}.

The new variant of the Seidelian iteration gives the following opening values, down and then up, after which we continue with vector-differences in phase:

2.2634									
1.5	1.6432	1.6579	1579	417	285	98	86	37	1.6987
1.5	1.4931	1.4453	-547	-643	-317	-209	-102	-48	1.3986
1.0	1.1372		1372		442		135	59	1.2008

For example, $1.5 + 0.1579 + 0.0285 + 0.0086 + 0.0037 = 1.6987$.

Here there is little to choose between the old and the new Seidelian iteration, for it so happens that the largest characteristic root of the former is real and fairly small.

Simple iteration, on the other hand, when performed with the change of matrix origin mentioned in § 3, so that for example the first equation of iteration is

$$6.34x_1^{(1)} = 3.17x_1^{(0)} - 0.92x_2^{(0)} + 1.07x_3^{(0)} - 1.13x_4^{(0)},$$

converges, but so much more slowly than the Seidelian iterations that it seems unnecessary to give the details. In particular, if we begin with the initial values {2.0 1.5 1.5 1.0}, the successive increments for x_1 take a long time in settling down to an approximate geometrical progression, of common ratio about 0.75.

REFERENCES TO LITERATURE

- AITKEN, A. C., 1925. "On Bernoulli's Numerical Solution of Algebraic Equations", *Proc. Roy. Soc. Edin.*, XLVI, 289-305.
- , 1937. "The Evaluation of the Latent Roots and Latent Vectors of a Matrix", *Proc. Roy. Soc. Edin.*, LVII, 269-304.
- BODEWIG, E., 1947. "Bericht über die verschiedenen Methoden zur Lösung eines System linearer Gleichungen mit reellen Koeffizienten", *Indag. Math.*, IX, 441-452, 1104-1116, 1285-1295; X, 53-64, 211-219.
- FRAZER, R. A., DUNCAN, W. J., and COLLAR, A. R., 1938. *Elementary Matrices*, Cambridge University Press.
- STEFFENSEN, J. F., 1933. "Remarks on Iteration", *Skandinavisk Aktuarietidskr.*, 64-72.
- TURNBULL, H. W., and AITKEN, A. C., 1932. *Theory of Canonical Matrices*, Blackie & Son, London and Glasgow.
- WHITTAKER, E. T., and ROBINSON, G., 1924. *The Calculus of Observations*, Blackie & Son, London and Glasgow.

(Issued separately June 16, 1950)

V.—Les transformations asymptotiquement presque périodiques discontinues et le lemme ergodique. (Première Note.) Par **Maurice Fréchet**, Hon.F.R.S.E., Université de Paris, à la Sorbonne. *Communicated by Sir EDMUND WHITTAKER*, F.R.S.

(MS. received October 4, 1948. Revised MS. received April 8, 1949. Read May 2, 1949)

SYNOPSIS

With the aim of establishing, under wide conditions, the ergodic theorem of G. D. Birkhoff, the author extends the class of asymptotically almost-periodic functions, considering now not only continuous functions, as he had already done in 1943, but discontinuous functions. Definitions and properties of the extended class of functions are set out, some comparisons being made with almost-periodic functions in the sense of Bohr, Stepanoff, Weyl and Besicovitch. Applications to the ergodic theorem are adumbrated.

INTRODUCTION

NOUS avons montré précédemment * comment l'introduction des fonctions A.P.P.C. (asymptotiquement presque périodiques continues) permet de simplifier considérablement—et par une méthode applicable en bien d'autres questions—la démonstration du lemme ergodique de George Birkhoff. En montrant que la fonction $f(T, M)$ qui y intervient appartient à la classe des fonctions A.P.P.C. beaucoup plus restreinte que celle des fonctions qui ont une moyenne à l'infini, notre méthode fournit en outre un résultat beaucoup plus précis que le dit lemme. Elle permet enfin d'étendre la validité de ce lemme dans une première direction en supprimant plusieurs des hypothèses sur lesquelles était fondé ce lemme. Il est vrai qu'en même temps notre méthode en restreignait la validité dans une seconde direction, en introduisant certaines hypothèses de continuité non exigées par ce lemme.

Même si l'on s'en tenait là, notre méthode fournissait des résultats nouveaux. Et même si elle n'avait apporté qu'une autre forme de démonstration, elle aurait conservé quelque intérêt.

Mais nous avons déjà, à cette époque, indiqué que nos hypothèses de continuité ne jouaient pas un rôle essentiel, qu'elles rendaient seulement plus simple la préparation à la démonstration, préparation consistant dans

* "Les fonctions asymptotiquement presque périodiques et leur application au problème ergodique", *Revue Scientifique*, 79^{ème} Année, 1941, pp. 341-354 et 407-417.

l'étude des propriétés, utiles en elles-mêmes, des fonctions A.P.P.C. Nous avons même pensé pouvoir étendre (aux pages 416-417 de notre Mémoire), notre démonstration au cas discontinu. Mais notre extension se basait sur un énoncé de W. H. Young que nous avons reproduit de mémoire inexactement comme nous l'a fait observer M. Ky Fan. L'extension ne pouvait donc être considérée comme prouvée ainsi que nous l'avons fait observer dans un mémoire ultérieur.*

Nous allons donc reprendre la question et montrer comment on peut généraliser les fonctions A.P.P.C. et se servir de fonctions asymptotiquement presque périodiques mais discontinues pour étendre la validité de notre démonstration. Il suffit pour cela d'appliquer à une telle généralisation l'esprit des indications qui se trouvaient déjà aux pages 343 et 344 de notre premier Mémoire cité ci-dessus.

Nous commencerons à nous placer dans un cas très général que nous particulariserons ensuite quand ce sera nécessaire ou au moins plus commode.

TRANSFORMATIONS ASYMPTOTIQUEMENT PRESQUE PÉRIODIQUES. CONVERGENCE AU SENS L_γ

Considérons une famille \mathcal{F} de transformations $N = \Phi(t)$ d'un nombre réel t en un élément N d'un ensemble E d'éléments de nature quelconque, chacune de ces transformations n'étant d'ailleurs nécessairement définie que sur une "demi-droite positive" $t > \alpha$, \dagger éventuellement variable avec la transformation. Supposons en outre, qu'en considérant \mathcal{F} comme support d'un espace fonctionnel \mathcal{E} dont chaque $\Phi(t)$ est un point, on traite \mathcal{E} comme un "espace L ", c'est à dire un espace où l'on ait défini la limite d'une suite de tels "points". Plus précisément, à chaque demi-droite positive $t > \gamma$, on associe une définition L_γ de la limite, sur cette demi-droite, d'une suite d'éléments $\Phi_n(t)$ de \mathcal{F} vers un élément $\Phi(t)$ de \mathcal{F} . (On suppose: 1° que si sur $t > \gamma$, Φ_n tend vers Φ au sens L_γ , il en est de même de toute suite extraite de la suite des Φ_n et 2° que si les fonctions $\psi_n(t)$ sont identiques sur $t > \gamma$ à une fonction $\psi(t)$ de \mathcal{F} , alors elles convergent vers ψ au sens L_γ .) On suppose, en outre, les L_γ telles que si $\gamma' > \gamma$, toute suite $\Phi_n(t)$ convergeant vers $\Phi(t)$ sur $t > \gamma$ au sens L_γ , converge aussi vers $\Phi(t)$ sur $t > \gamma'$ au sens $L_{\gamma'}$. Par exemple, en prenant pour E l'ensemble des nombres réels, on pourrait prendre pour L_γ la convergence presque partout sur $t > \gamma$, de sorte qu'il n'est pas nécessaire de supposer $\Phi(t)$ continue.

* "Sur le problème ergodique", *Revue Scientifique*, 81^e Année, 1943, pp. 115-157.

† Voir le premier article cité ci-dessus, pour les raisons qui nous font considérer ici une demi-droite au lieu d'une droite entière et un peu plus loin une certaine limite quand un nombre K tend vers $+\infty$ au lieu de $\pm\infty$.

TRANSFORMATIONS ASYMPTOTIQUEMENT PRESQUE PÉRIODIQUES A DROITE AU SENS L

Ceci étant, nous dirons qu'une transformation $F(t)$ de \mathcal{F} est asymptotiquement presque périodique à droite* au sens L , si quelle que soit la suite σ de nombres h_n tendant vers $+\infty$, il existe une suite de nombres k_n , extraite de σ et une fonction $p(t)$ de \mathcal{F} telles que, quel que soit le nombre γ , $F(t + k_n)$ converge vers $p(t)$ au sens L_γ sur $t > \gamma$. (Ceci suppose donc implicitement que $p(t)$ est défini quel que soit le nombre t .)

Définition équivalente.—Il est d'ailleurs possible de mettre cette définition sous une forme équivalente (quoique en apparence moins stricte), et surtout plus commode pour certaines démonstrations. Pour qu'une transformation $N = F(t)$ de \mathcal{F} soit asymptotiquement presque périodique au sens L , il faut et il suffit que, quel que soit le nombre α et quelle que soit la suite σ des $h_n \rightarrow \infty$, il existe une suite de nombres $k_n^{(\alpha)}$ extraite de σ telle que la suite des $F(t + k_n^{(\alpha)})$ soit convergente au sens L_α sur $t > \alpha$.

La condition est évidemment nécessaire. Pour prouver qu'elle est suffisante, utilisons le procédé de la diagonale. Si la condition est remplie, appliquons-la pour $\alpha = -1$; posons $k_n^{(1)} = k_n^{(-1)}$ et soit $p^{(-1)}(t)$ la limite au sens L_{-1} des $F(t + k_n^{(1)})$. Appliquons encore la condition pour $\alpha = -2$, mais en remplaçant les h_n par les $k_n^{(1)}$; on en extraira une suite de nombres $k_n^{(2)}$ et la suite des $F(t + k_n^{(2)})$ convergera au sens L_{-2} vers une transformation $p^{(-2)}(t)$. D'après ce qui précède $p^{(-1)}(t)$ coïncide avec $p^{(-2)}(t)$ pour $t > -1$. On continuera ainsi et on formera des suites de nombres $k_n^{(r)}$ extraits de la suite des $k_n^{(r-1)}$ et tels que les $F(t + k_n^{(r)})$ convergeront au sens L_{-r} sur $t > -r$ vers une transformation $p^{(-r)}(t)$. Soit alors $k_n = k_n^{(n)}$; la suite des k_n est extraite de la suite des k_n . Et comme à partir du rang r , elle est aussi extraite de la suite des $k_n^{(r)}$, la suite $F(t + k_n)$ converge au sens L_{-r} vers $p^{(-r)}(t)$ pour $t > -r$. D'ailleurs, appelons $p(t)$ la fonction qui coïncide avec $p^{(-r)}(t)$ pour $-r \leq t < -r+1$ et avec $p^{(-1)}(t)$ pour $t > -1$; elle coïncide alors aussi pour toute valeur de $t > -r$ avec $p^{(-r)}(t)$. Soit maintenant α un nombre quelconque; il y a au moins un entier positif r tel que $-r < \alpha$; alors $F(t + k_n)$ convergeant au sens L_{-r} vers $p^{(-r)}(t)$ sur $t > -r$, converge au même sens vers $p(t)$ sur $t > -r$ et par suite converge au sens L_α vers $p(t)$ sur $t > \alpha$: La condition admise est donc bien suffisante.

Remarque.—La définition que nous venons d'adopter est la plus commode pour le cas actuel. Mais, au point de vue intuitif, elle a l'inconvénient de ne rien faire apparaître qui justifie l'expression "presque périodique". En particulierisant ce qui précède au cas où la limite au sens L peut être exprimée au moyen d'une distance, on peut, comme nous le ferons ailleurs, prouver l'existence des "presque-périodes". Et l'hypothèse d'une distance permettra même d'ajouter des propriétés supplémentaires à celles qui seront établies ici.

* Pour abrégé, nous supprimerons dans la suite les mots "à droite".

Généralisation.—Soit $\psi(t) = U[\Phi(t)]$ une opération fonctionnelle faisant correspondre à des transformations $N = \Phi(t)$ de la famille \mathcal{F} (où l'on a défini la limite au sens L) des transformations $P = \psi(t)$ appartenant à une famille \mathcal{F}' (où l'on a défini la limite en un certain sens L'), P appartenant à un ensemble E' d'éléments de nature quelconque (E' pouvant être ou non distinct de E). Nous supposons de plus que, pour toute constante positive h , on a $U[\Phi(t+h)] = \psi(t+h)$.*

Condition D.—On dira que l'opération U est *continue* si, quand $\Phi_n(t)$ de \mathcal{F} tend au sens L vers $\Phi(t)$ de \mathcal{F} , $U[\Phi_n(t)]$ de \mathcal{F}' tend au sens L' vers $U[\Phi(t)]$ de \mathcal{F}' . Quand il en est ainsi, l'opération U transforme alors toute transformation $\Phi(t)$ qui est asymptotiquement presque périodique au sens L en une transformation $\psi(t)$ qui est asymptotiquement presque périodique au sens L' . Car si $h_n \rightarrow +\infty$, on peut en tirer une suite k_1, k_2, \dots telle que $\Phi(t+k_n)$ converge au sens L vers une fonction $p(t)$ de \mathcal{F} et alors $\psi(t+k_n) = U[\Phi(t+k_n)]$ converge au sens L' vers une fonction $U[p(t)]$ de \mathcal{F}' .

En particulier, soit $P = f(N)$ une transformation de points N de E en points P de E' ; $f(\Phi(t))$ est une transformation de nombres t en points de E' et par suite $f(\Phi(t))$ définit une opération $U[\Phi(t)]$ d'une espèce particulière. Il se peut qu'elle soit continue au sens indiqué plus haut et alors si $\Phi(t)$ est asymptotiquement presque périodique au sens L , $f(\Phi(t))$ le sera aussi au sens L' .

APPLICATIONS AUX FAMILLES DE TRANSFORMATIONS PONCTUELLES

Soit d'abord S un ensemble borné et fermé de points d'un espace euclidien ou plus généralement un ensemble compact en soi † S formé de points d'un espace topologique E où une limite est définie, satisfaisant aux conditions 1°, 2° de la p. 62. Considérons maintenant une transformation $N = T_t(M)$ d'un point arbitraire M de S en un point correspondant N de S , cette transformation dépendant du paramètre numérique t supposé essentiellement > 0 . Nous ferons les hypothèses suivantes:—

(A) Pour chaque M fixe de S , $T_t(M)$ appartient à la famille \mathcal{F} .

(B) On peut alors considérer aussi $T_t(M)$ comme un point de l'espace fonctionnel \mathcal{E} , point dépendant du point M . Nous supposons que c'est une transformation continue au sens L , par rapport au point M de S .

* C'est, par exemple, ce qui a lieu quand les transformations $\Phi(t)$ sont des fonctions numériques, chacune absolument intégrable sur son domaine de définition et qu'on prend

$U[\Phi(t)] = \int_0^{+\infty} \Phi(x) dx$. C'est aussi ce qui a lieu dans l'exemple considéré plus loin, où $U[\Phi(t)] = f(\Phi(t))$.

† Un ensemble S est compact en soi lorsque dans toute suite de points M_n de S , il existe une sous-suite de points M_{n_1}, M_{n_2}, \dots qui converge vers un point de S .

C'est à dire que si M_n de S tend vers M de S , $T_t(M_n)$ considéré comme fonction de t converge vers $T_t(M)$ au sens L_0 sur $t > 0$. Nous supposons en outre que quel que soit h , $T_{t+h}(M_n)$ converge vers $T_{t+h}(M)$ au sens L_{-h} .

(C) On a

$$T_{t+t'}(M) = T_t(T_{t'}M)$$

pour tout couple $t > 0$, $t' > 0$.

Sous ces trois hypothèses, $T_t(M)$ est une transformation asymptotiquement presque périodique au sens L . En effet, soit h_1, h_2, \dots une suite de nombres tendant vers $+\infty$, et α un nombre quelconque. On peut trouver r tel que $h_r + \alpha > 0$ et $h_r > 0$ et trouver ν tel que $h_n - h_r > 0$ pour $n > \nu$. On peut alors écrire pour $n > \nu$ et $t > \alpha$

$$T_{t+h_n}(M) = T_{t+h_r}(T_{h_n-h_r}(M)).$$

La suite des points $T_{h_n-h_r}(M)$ appartenant à S contient une sous-suite convergente de points $T_{h_n-h_r}(M)$; soit $M_\infty^{(r)}$ sa limite qui appartient à S . Pour $t > \alpha$, $t + h_r$ sera > 0 et la suite de transformations $T_{t+h_n}(M)$ convergera au sens L_α vers la fonction $T_{t+h_r}(M_\infty^{(r)})$.

Revenons maintenant à l'opération U considérée ci-dessus, faisant correspondre à une transformation $N = F(t)$ de \mathcal{F} une transformation $P = \psi(t)$ de \mathcal{F}' et vérifiant encore la condition

$$(K) \quad U[F(t+h)] = \psi(t+h).$$

Si cette opération est continue au sens indiqué plus haut, alors $U[T_t M]$ est une transformation appartenant à \mathcal{F}' et qui, dans les hypothèses (A), (B), (C), (K) sur $T_t M$, sera aussi une transformation asymptotiquement presque périodique au sens L' .

Moyenne à l'infini.—On sait que les fonctions presque périodiques, au sens de Bohr,—qui sont continues—, ou aux sens de Stepanoff, de H. Weyl ou de Besicovitch,*—qui peuvent être discontinues—, ont une moyenne à l'infini.† Il en est donc de même des fonctions asymptotiquement presque périodiques correspondantes. D'autre part, sur les quatre définitions correspondantes, les trois premières sont des cas particuliers‡ de la nôtre ci-dessus. Nous allons faire une hypothèse supplémentaire qui peut donc sûrement être vérifiée—et qui l'est même dans trois cas qui sont parmi les plus simples et les plus importants—en nous plaçant maintenant dans le cas où :

* Voir pour leurs définitions, *Leçons sur les fonctions presque périodiques* par Favard. Gauthier-Villars, 1933.

† Définition plus loin, p. 66.

‡ A. S. Voir Kovanko, "Sur la compacité des systèmes de fonctions presque périodiques généralisées de H. Weyl", *C.R. Acad. Sc. U.R.S.S.*, XLIII, 1944, 275-276.

(H) La définition de la limite L' dans la famille \mathcal{F}' est telle que toute transformation asymptotiquement presque périodique au sens L' , soit $P = \psi(t)$, a une moyenne à l'infini.

Cela suppose d'abord que la définition de la moyenne à l'infini peut avoir un sens dans le cas où E et E' sont de nature quelconque.

Nous dirons que $\psi(t)$ de \mathcal{F}' a une moyenne à l'infini (à droite), quand il existe au moins un nombre β (sur la demi-droite $t > \alpha$ où $\psi(t)$ est défini) tel que le rapport

$$\frac{1}{K} \int_{\beta}^{\beta+K} \psi(t) dt \quad (1)$$

tende vers une limite déterminée quand $K \rightarrow +\infty$. Comme ici $N = \psi(t)$ est un point de l'ensemble E' , dont jusqu'ici nous avons supposé les éléments de nature quelconque, il faut d'abord donner un sens à l'intégrale et à la limite. On peut lui en donner un quand E' est un espace vectoriel distancié, dit aussi de Banach-Wiener.

Pour cet espace, la généralisation de l'intégrale classique au sens de Cauchy est immédiate, comme limite unique, si elle existe, de sommes de la forme classique

$$\sum_{i=1}^n \Phi(\xi_i)(t_i - t_{i-1}).$$

Mais il est plus utile de généraliser l'intégrale de Stieltjes-Lebesgue en appliquant au cas actuel la définition plus générale encore que nous avons donnée ailleurs.* Il nous suffira de retenir que, quand cette intégrale existe,

(1°) elle est la limite unique quand $\epsilon \rightarrow 0$ de sommes de la forme

$$s = \sum_i \Phi(\xi_i) \cdot (\text{mesure } e_i)$$

où l'on a décomposé $(\beta, \beta + K)$ en une suite finie ou dénombrable d'ensembles mesurables disjoints e_i , dans chacun desquels "l'oscillation" de $\Phi(t)$ est $< \epsilon$, où ξ_i est un point arbitraire de e_i et où s est supposé absolument convergente;

(2°) cette intégrale possède celles des propriétés classiques de l'intégrale ordinaire que nous aurons à utiliser dans la suite.

* "L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque", *Revue Scientifique*, 82^e Année, 1944, pp. 483-512.

Cas cartésien.—Dans le cas où E est un espace cartésien à un nombre fini, r , de dimensions, tout point de E peut être défini par une suite ordonnée d'un nombre fini fixe, r , de nombres appelés coordonnées de ce point. Alors on voit que le rapport (1) représentera le point de E dont les coordonnées sont

$$\frac{1}{K} \int_{\beta}^{\beta+K} \psi_k(t) dt \quad k=1, \dots, r \quad (2)$$

où $\psi_k(t)$ est la k^{e} coordonnée de $\psi(t)$.*

Nous voyons que $\psi(t)$ a une moyenne à l'infini si les rapports (2) tendent vers des limites respectives quand $K \rightarrow \infty$, limites que nous désignerons par $\mathfrak{M}_{+\infty} \psi_k(t)$. Ainsi, la moyenne à l'infini du point mobile $\psi(t)$ de E sera le point de E , désigné par $\mathfrak{M}_{+\infty} \psi(t)$, qui a pour coordonnées les $\mathfrak{M}_{+\infty} \psi_k(t)$ et on pourra écrire

$$\mathfrak{M}_{+\infty} \psi(t) = \lim_{K \rightarrow +\infty} \frac{1}{K} \int_{\beta}^{\beta+K} \psi(t) dt \quad (3)$$

Cas de l'espace de Banach-Wiener.—La même notation sera employée quand E' est un espace vectoriel distancié. Il n'est pas nécessaire de faire intervenir dans le symbole du premier membre la quantité β qui figure au second membre. Car si la limite (3) existe pour une valeur de β (telle que $\psi(t)$ soit défini au moins pour $t > \beta$) elle existe aussi (et a la même valeur) pour toute autre valeur $\gamma (> a)$ de β . On peut en effet écrire

$$\frac{1}{K} \int_{\gamma}^{\gamma+K} \psi(t) dt = \frac{1}{K} \int_{\gamma}^{\beta} \psi(t) dt + \frac{\gamma - \beta + K}{K} \cdot \frac{1}{\gamma - \beta + K} \int_{\beta}^{\gamma+K} \psi(t) dt,$$

et le second membre tend évidemment vers le même point de E' que le second membre de (3).

Remarque.—On appelle souvent $\mathfrak{M}_{+\infty} \psi(t)$: moyenne de $\psi(t)$. Nous préférons l'appeler moyenne à l'infini (à droite) pour bien marquer ce fait—qui résulte de ce qui précède—que l'existence et la valeur de cette moyenne ne sont pas modifiées quand on remplace $\psi(t)$ sur un segment fini quelconque par une autre transformation également quelconque, pourvu qu'elle soit, comme $\psi(t)$ intégrable sur ce segment. Autrement dit, l'existence et la valeur de $\mathfrak{M}_{+\infty} \psi(t)$ ne dépendent—si l'on peut s'exprimer ainsi—que des valeurs de $\psi(t)$ à l'infini (à droite).

Application au lemme ergodique.—Tout ceci étant, il résulte de ce qui précède que si les définitions de la limite aux sens L et L' , celle de la transformation $T_t M$ et celle de l'opération U , satisfont aux conditions

* Ce langage suppose implicitement que $\psi(t)$ soit intégrable sur tout segment fini de $t > a$, (c. a. d. qu'il en soit ainsi pour ses coordonnées).

(A), (B), (C), (K), (H), alors $U[T, M]$ a une moyenne à l'infini. Mais nous avons, dans ce qui précède, démontré *plus* que cette propriété (qui constitue le résultat du lemme ergodique), puisque nous avons même prouvé que $U[T, M]$ est une fonction de t qui est asymptotiquement presque périodique au sens L' .

Remarque.—Dans un mémoire ultérieur, nous indiquerons les propriétés supplémentaires qu'on obtient dans le cas où les limites L, L' peuvent être définies par l'intermédiaire d'une distance et enfin nous appliquerons ces résultats au cas spécial considéré par George Birkhoff.

(Issued separately June 16, 1950)

VI.—Unbiased Statistics with Minimum Variance. By **A. Bhattacharyya**, Statistical Laboratory, Calcutta. *Communicated by Professor A. C. AITKEN, F.R.S.*

(MS. received April 26, 1949. Read July 4, 1949)

SYNOPSIS

The problem considered is that of the estimation of a statistical parameter from a sample of values of the variate or variates concerned. Reference is made to the method of unbiased statistics with minimum variance, developed by Aitken and Silverstone. The principal result obtained by these authors is generalized, and an inequality involving the variances of unbiased statistics is obtained. Several examples illustrating the theory are appended.

1. THE problem of estimation has been approached from various angles. Of these, perhaps the earliest was that of "unbiased statistics with minimum variance". The method of least squares of Gauss and that of linear estimation of Markoff were based on this concept. Fisher (1921) laid the foundation of his theory of estimation on this concept—his criteria of consistency and efficiency are nothing but large sample set-up of that of unbiasedness and minimum variability. Aitken revived the problem again and, in collaboration with Silverstone (1941), made an attempt to arrive at some general solution. The last-named author showed, with the help of the Calculus of Variations, that under some regularity conditions and in the case when the ranges of the stochastic variables are independent of the parameters, a statistic $t_j(x_1, x_2, \dots, x_n)$ is an unbiased estimate with minimum variance of the parameter θ_j if

$$t_j = \theta_j - \lambda_j \frac{\partial \log F}{\partial \theta_j},$$

where $F(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_t) dx_1 dx_2 \dots dx_n$ is the joint probability distribution of the stochastic variables, x_1, x_2, \dots, x_n , $t_j(x_1, x_2, \dots, x_n)$ is a function of these stochastic variables alone (free from the parameters $\theta_1, \theta_2, \dots, \theta_t$), and $\lambda_j(\theta_1, \theta_2, \dots, \theta_t)$ is a function of the parameters $\theta_1, \theta_2, \dots, \theta_t$ only. The author in a series of papers (1946–48) obtained a more general result, though the method followed by him in proving this was different from that of Aitken and Silverstone. It was shown that, under some general conditions similar to those stated

above, the statistic $t(x_1, x_2, \dots, x_n)$ is an unbiased estimate with minimum variance of the function $\tau(\theta_1, \theta_2, \dots, \theta_l)$ of the parameter if

$$t = \tau + \frac{1}{F} \sum \lambda_{i_1, i_2, \dots, i_l} \frac{\partial^{i_1+i_2+\dots+i_l} \tau}{\partial \theta_1^{i_1} \partial \theta_2^{i_2} \dots \partial \theta_l^{i_l}}, \quad (1.1)$$

where the summation is over the i 's, and the λ 's are functions of the parameters θ 's. It was further shown that in case a statistic satisfying (1.1) did not exist, even then it was possible to derive a lower bound of the variances of unbiased statistics estimating τ . It will be readily seen that the result of Aitken and Silverstone is a special case of this general result. We now propose to generalize the result (1.1) still further and obtain a more extended result. In deriving this we shall follow the method of the Calculus of Variations, though the method already used by the author in the previous papers (namely, the method of regression) is also available and in some respects more suitable. Thus, it should be remarked that a result deduced by the method of the Calculus of Variations is only valid when the (stochastic) variables are continuous, whereas the result derived by us is valid both for continuous and discrete cases (as can be shown by deriving this by the regression method).

2. Let the n stochastic variables x_1, x_2, \dots, x_n (independent or dependent) follow the probability distribution

$$F(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_l) dx_1 dx_2 \dots dx_n, \quad (2.1)$$

where $\theta_1, \theta_2, \dots, \theta_l$ are l parameters involved in the distribution. Let $\tau(\theta_1, \theta_2, \dots, \theta_l)$ be a given function of the parameters. The problem is to find a function of the stochastic variables $t(x_1, x_2, \dots, x_n)$ such that its expectation is the given function τ and its variance is a minimum. Analytically we have to find the function t such that

$$\int (t - \tau)^2 F dv = \text{a minimum},$$

subject to the condition

$$\int t F dv = \tau, \quad (2.2)$$

where for brevity dv has been written for the elementary volume $dx_1 dx_2 \dots dx_n$, and the integral extends over all the variables over their entire range. As τ is supposed to be given, our purpose is served if we can find a t such that

$$\int t^2 F dv = \text{a minimum}, \quad (2.3)$$

subject to the same condition (2.2). The problem as stated in the general

form has not been solved yet, and there may not be a general solution at all for all distributions. It may happen that very often there may not exist an unbiased statistic which satisfies (2.2), and in that case it is futile and irrelevant to proceed in this way. Or it may happen that in a particular distribution only some classes of functions of the parameters have unbiased estimates. Even then the following method may not lead to a statistic with minimum variance. In short, we do not claim to have found a necessary condition under which an unbiased statistic with minimum variance exists; what we have found is a sufficient condition which, if satisfied, leads to the determination of an unbiased statistic with minimum variance (provided it exists).

To get a practical solution we assume that the limits of integration in (2.2) or (2.3) are independent of the parameters $\theta_1, \theta_2, \dots, \theta_l$. This assumption regarding the independence of the range of the stochastic variables of the parameters severely limits the applicability of this method; but this restriction was assumed by Aitken and Silverstone also. Then, under some regularity conditions (*e.g.* differentiability of the frequency function F with respect to the parameters θ 's, uniform convergence of the integrals involving such functions, etc.), we can derive other relations from (2.2) with the help of integration and differentiation (including repeated process) with respect to the parameters $\theta_1, \theta_2, \dots, \theta_l$. For example, we can deduce the following relations:—

$$\begin{aligned}
 \text{(i)} \quad & \int t \frac{\partial^{t_1+t_2+\dots+t_l}}{\partial \theta_1^{t_1} \partial \theta_2^{t_2} \dots \partial \theta_l^{t_l}} F, & dv &= \frac{\partial^{t_1+t_2+\dots+t_l}}{\partial \theta_1^{t_1} \partial \theta_2^{t_2} \dots \partial \theta_l^{t_l}} \tau; \\
 \text{(ii)} \quad & \int t \left[\frac{\partial^r}{\partial \theta_i^r} \int_{\theta_i^0}^{\theta_i^1} h(\theta_1, \theta_2, \dots, \theta_l) F d\theta_i \right], & dv &= \frac{\partial^r}{\partial \theta_i^r} \int_{\theta_i^0}^{\theta_i^1} h(\theta_1, \theta_2, \dots, \theta_l) \tau d\theta_i; \\
 \text{(iii)} \quad & \int t \left[\int_{\theta_i^0}^{\theta_i^1} \int_{\theta_j^0}^{\theta_j^1} h(\theta_1, \theta_2, \dots, \theta_l) F d\theta_i d\theta_j \right], & dv &= \int_{\theta_i^0}^{\theta_i^1} \int_{\theta_j^0}^{\theta_j^1} h(\theta_1, \theta_2, \dots, \theta_l) \tau d\theta_i d\theta_j,
 \end{aligned}$$

where θ_i^0, θ_i^1 are suitable constants, and $h(\theta_1, \theta_2, \dots, \theta_l)$ is some suitable function of the parameters $\theta_1, \theta_2, \dots, \theta_l$.

The above are given by way of illustration, and the reader can easily construct more or less complicated examples. But these or similar expressions are generally of the form

$$\int t(x_1, x_2, \dots, x_n) \Phi(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_l) dv = \alpha(\theta_1, \theta_2, \dots, \theta_l), \quad (2.4)$$

when the Φ 's are some functions of the variables x_1, x_2, \dots, x_n and the parameters $\theta_1, \theta_2, \dots, \theta_l$, which can be derived from $F(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_l)$, and the α 's are some functions derivable from

$\tau(\theta_1, \theta_2, \dots, \theta_i)$ without any prior knowledge of the actual statistic t . Thus in the above cases we have

$$\begin{aligned} \text{(i)} \quad \Phi &= \frac{\partial^{t_1+t_2+\dots+t_i}}{\partial \theta_1^{t_1} \partial \theta_2^{t_2} \dots \partial \theta_i^{t_i}} F, & \alpha &= \frac{\partial^{t_1+t_2+\dots+t_i}}{\partial \theta_1^{t_1} \partial \theta_2^{t_2} \dots \partial \theta_i^{t_i}} \tau; \\ \text{(ii)} \quad \Phi &= \frac{\partial^r}{\partial \theta_i^r} \int_{\theta_j}^{\theta_j'} h(\theta_1, \theta_2, \dots, \theta_i) F d\theta_i, & \alpha &= \frac{\partial^r}{\partial \theta_i^r} \int_{\theta_j}^{\theta_j'} h(\theta_1, \theta_2, \dots, \theta_i) \tau d\theta_i; \\ \text{(iii)} \quad \Phi &= \int_{\theta_j}^{\theta_j'} \int_{\theta_i}^{\theta_i'} h(\theta_1, \theta_2, \dots, \theta_i) F d\theta_i d\theta_j, & \alpha &= \int_{\theta_j}^{\theta_j'} \int_{\theta_i}^{\theta_i'} h(\theta_1, \theta_2, \dots, \theta_i) \tau d\theta_i d\theta_j. \end{aligned}$$

We may mention in passing that besides the use of the infinitesimal calculus, other calculus (*e.g.* finite calculus) may be used to derive relations of this nature.

Now we are faced with the problem of finding a statistic $t(x_1, x_2, \dots, x_n)$ such that the integral (2.3) is a minimum subject to the restriction (2.2) and all restrictions of the nature (2.4) derivable from it. Introducing Lagrangian undetermined multipliers λ 's, we can reduce the problem to one of unrestricted maximum or minimum. Thus, multiplying each of the relations (2.4) and (2.2) by means of 2λ and subtracting this from (2.3), we get

$$\int [tF - \lambda_0 tF - 2t \sum \lambda_i \Phi_i] dv$$

which has to be minimized. The Euler equation is obtained partially differentiating the integrand with respect to t and equating that to zero. Thus we get

$$tF - \lambda_0 F - \sum \lambda_i \Phi_i = 0 \quad (2.5a)$$

or

$$t = \lambda_0 + \frac{1}{F} \sum \lambda_i \Phi_i. \quad (2.5)$$

Thus t is determined by (2.3). We get the interesting result that if in any distribution an expression of the nature of one given on the right side of (2.5) is free from the population parameters, then this gives an unbiased statistic with minimum variance for the estimation of its own expectation (provided the variance itself exists). To determine the λ 's we can make use of the relations (2.4) and (2.2). Thus multiplying (2.5) by Φ_j and integrating over the x 's, we get

$$\alpha_j = \int t \Phi_j dv - \lambda_0 \int \Phi_j dv + \sum \lambda_i \int (\Phi_i \Phi_j / F) dv. \quad (2.6)$$

The integrals on the right-hand side involve the Φ 's which are all known,

and so they themselves are uniquely known. We here do not enter into the question of convergence of these integrals, which are all assumed to be convergent. Incidentally we remark that this restricts the choice of the k 's in the illustrated cases of Φ 's given above.

The set of linear equations (2.6) involving the λ 's gives us their values on being solved (the solvability of these linear equations is taken for granted). The variance of such a statistic, when it exists, is easily found. Thus from (2.5) we get, after multiplying both sides by t and integrating,

$$\int t^2 F dv = \lambda_0 \tau + \sum \lambda_i \int t \Phi_i dv = \lambda_0 \tau + \sum \lambda_i a_i,$$

and so the variance of t is

$$V(t) = (\lambda_0 - \tau)\tau + \sum \lambda_i a_i. \quad (2.7)$$

3. The method followed above involved the use of the Calculus of Variations; so it is implied that the stochastic variables are continuous, or satisfy similar conditions under which the Calculus of Variations is applicable. Also, the theory of the Calculus of Variations is very abstract in nature and so it is very difficult to comprehend the meaning and scope of the result (2.5). The regression method used by the author in the previous paper is more simple and straightforward. Besides, this latter method is applicable for discrete or continuous stochastic variables. We can use this method to deduce the result (2.5) and thereby prove that the relation (2.5) gives a sufficient condition for the existence of an unbiased statistic with minimum variance.

We have already remarked that, in the distribution, any function of the parameters may not possess an unbiased statistic. We have just now observed that a statistic which is of the form (2.5) may not exist, although there may be unbiased statistics. In the second case, though we do not get a decisive answer to our search for the unbiased statistic with minimum variance, a very interesting result can be obtained by the regression method about the lower bound of the variances of unbiased statistics. As this method has been discussed in the previous papers, we give the result without the proof:

$$V(T) > (\lambda_0 - \tau)\tau + \sum \lambda_i a_i,$$

where T is a statistic whose expectation is τ . The sign of equality holds only when T is of the form (2.5), when it yields the unbiased statistic with minimum variance.

4. Some simple examples are considered below to illustrate the above theory.

A. The *binomial* distribution is first considered as an example of a uni-parametric discrete distribution. Here

$$F(x, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}.$$

So we have the following results (which can be easily verified)

$$(i) \quad \frac{x}{n} = p + \frac{p(1-p)}{n} \frac{1}{F} \frac{\partial F}{\partial p},$$

which shows that x/n is an unbiased statistic for the estimation of p with minimum variance.

$$(ii) \quad \frac{\frac{x}{n} \left(1 - \frac{x}{n}\right)}{n-1} = \frac{p(1-p)}{n} + \frac{p(1-p)(1-2p)}{n^2} \frac{1}{F} \frac{\partial F}{\partial p} - \frac{p^2(1-p)^2}{n^2(n-1)} \frac{1}{F} \frac{\partial^2 F}{\partial p^2},$$

which shows that $\frac{x}{n} \left(1 - \frac{x}{n}\right) / (n-1)$ is an unbiased statistic with minimum variance for the estimation of $p(1-p)/n$ (which is the variance of the estimate of p).

$$(iii) \quad \frac{1}{x+r} = \frac{(1-p)^{n+r}}{p^r} \cdot \frac{1}{F} \int_0^p p^{r-1} (1-p)^{-n-r-1} F dp,$$

which shows that $1/(x+r)$ is an unbiased statistic with minimum variance. The mathematical expectation of this is

$$E\left(\frac{1}{x+r}\right) = \frac{(1-p)^{n+r}}{p^r} \int_0^p p^{r-1} (1-p)^{-n-r-1} dp,$$

which reduces to $[1 - (1-p)^{n+1}] / (n+1)p$ when $r=1$. This shows that $(n+1)/(x+1)$ is a statistic with minimum variance for the estimation of $1/p$ with a negative bias equal to $(1-p)^{n+1}/p$. This bias becomes negligible if n is sufficiently large or if p is very nearly equal to unity.

B. The next distribution to be considered is the two-parametric *normal* distribution. The distribution function of a sample from a normal distribution with population mean m and population standard deviation σ is

$$F(x_1, x_2, \dots, x_n; m, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{\sum (x_i - m)^2}{2\sigma^2} \right\}.$$

The following results can be easily verified:—

$$(i) \quad \frac{\sum x_i}{n} = m + \frac{\sigma^2}{n} \frac{1}{F} \frac{\partial F}{\partial m},$$

which shows that $\sum x_i/n = \bar{x}$ is an unbiased statistic with minimum variance for the estimation of the parameter m .

$$(ii) \quad \frac{\sum (x_i - \bar{x})^2}{n-1} = \sigma^2 + \frac{\sigma^4}{n-1} \frac{1}{F} \frac{\partial F}{\partial \sigma} - \frac{\sigma^4}{n(n-1)} \frac{1}{F} \frac{\partial^2 F}{\partial m^2},$$

which shows that $\sum (x_i - \bar{x})^2/(n-1)$ is an unbiased statistic with minimum variance for the estimation of σ^2 .

C. Two samples, x_1, x_2, \dots, x_n and $x'_1, x'_2, \dots, x'_{n'}$, of size n and n' , are taken from two normal populations with zero mean and standard deviations σ and σ' respectively. Then the distribution function of these two samples is

$$F(x_1, x_2, \dots, x_n; x'_1, x'_2, \dots, x'_{n'}; \sigma, \sigma') \\ = \frac{1}{(\sqrt{2\pi}\sigma)^n (\sqrt{2\pi}\sigma')^{n'}} \exp \left\{ -\frac{\sum x_i^2}{2\sigma^2} - \frac{\sum x'_i{}^2}{2\sigma'^2} \right\}.$$

It can easily be shown that

$$\sum x_i^2 / \sum x'_i{}^2 = \frac{1}{\sigma^{n-2} \sigma'^{n'}} \frac{1}{F} \int_0^{\sigma'} \frac{\partial}{\partial \sigma} (\sigma^n \sigma'^{n'-2} F) d\sigma',$$

and hence $\sum x_i^2 / \sum x'_i{}^2$ is an unbiased statistic with minimum variance. It is also seen that

$$E(\sum x_i^2 / \sum x'_i{}^2) = \frac{1}{\sigma^{n-2} \sigma'^{n'}} \int_0^{\sigma'} \frac{\partial}{\partial \sigma} (\sigma^n \sigma'^{n'-2}) d\sigma' = \frac{n}{n'-2} \frac{\sigma^2}{\sigma'^2},$$

and

$$E(\sum x_i^2 / \sum x'_i{}^2)^2 = \frac{1}{\sigma^{n-2} \sigma'^{n'}} \int_0^{\sigma'} \frac{\partial}{\partial \sigma} \left(\sigma^n \sigma'^{n'-2} \frac{n}{n'-2} \frac{\sigma^2}{\sigma'^2} \right) d\sigma' = \frac{n(n+2)}{(n'-2)(n'-4)} \frac{\sigma^4}{\sigma'^4}.$$

From these we can get the variance.

D. Let x and y be two correlated normal variables with zero mean, and let them have a distribution function of the form

$$(2\pi)^{-1} (a\beta - \gamma^2)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (ax^2 + \beta y^2 - 2\gamma xy) \right\}.$$

The distribution function of a sample of size n is of the form

$$F(x_1, y_1; x_2, y_2; \dots; x_n, y_n; a, \beta, \gamma) \\ = (2\pi)^{-n} (a\beta - \gamma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} (a\sum x_i^2 + \beta\sum y_i^2 - 2\gamma\sum x_i y_i) \right\}.$$

It is readily seen that

$$r^2 = (\Sigma x_i y_i)^2 / (\Sigma x_i^2)(\Sigma y_i^2) = \frac{(\alpha\beta - \gamma^2)^{\frac{n}{2}}}{4F} \frac{\partial^2}{\partial \gamma^2} \int_a^\infty \int_\beta^\infty (\alpha\beta - \gamma^2)^{-\frac{n}{2}} F d\alpha d\beta,$$

which shows that r^2 is a statistic with minimum variance. Now the mathematical expectation of r^2 is given by

$$\begin{aligned} E(r^2) &= \frac{1}{4} (\alpha\beta - \gamma^2)^{\frac{n}{2}} \frac{\partial^2}{\partial \gamma^2} \int_a^\infty \int_\beta^\infty (\alpha\beta - \gamma^2)^{-\frac{n}{2}} d\alpha d\beta \\ &= \frac{n}{4} (\alpha\beta - \gamma^2)^{\frac{n}{2}} \int_a^\infty \int_\beta^\infty (\alpha\beta - \gamma^2)^{-\frac{n}{2}-1} d\alpha d\beta \\ &\quad + \frac{n(n+2)}{4} \gamma^2 (\alpha\beta - \gamma^2)^{\frac{n}{2}} \int_a^\infty \int_\beta^\infty (\alpha\beta - \gamma^2)^{-\frac{n}{2}-2} d\alpha d\beta. \end{aligned}$$

It is possible to express these integrals in finite form, and so $E(r^2)$ can be expressed in a finite form. However, when n is large, this finite expression will contain a very large number of terms and so is not suitable. We can, however, find an approximate value in such cases. The behaviour of the integral of the form

$$I_N = \int_a^\infty \int_\beta^\infty (ab - \gamma^2)^{-N} d\alpha d\beta,$$

when N is large, can be studied by means of the transformation $a = \alpha + \frac{u(\alpha\beta - \gamma^2)}{N}$ and $b = \beta + \frac{v(\alpha\beta - \gamma^2)}{N}$, where u and v are the new variables. We get

$$I_N = \int_0^\infty \int_0^\infty [(\alpha\beta - \gamma^2) + u\beta(\alpha\beta - \gamma^2)/N + v\alpha(\alpha\beta - \gamma^2)/N + uv(\alpha\beta - \gamma^2)^2/N^2]^{-N} \frac{(\alpha\beta - \gamma^2)^2}{N^2} du dv.$$

Hence

$$N^2 (\alpha\beta - \gamma^2)^{N-2} I_N = \int_0^\infty \int_0^\infty [1 + u\beta/N + v\alpha/N + uv(\alpha\beta - \gamma^2)/N^2]^{-N} du dv.$$

Now as the integral on the right side is uniformly convergent for all values of α , β , and N , and as the integrand uniformly approaches the value $\exp(-u\beta - v\alpha)$ for all values of α , β , u , v as N tends to infinity, it is not difficult to prove that

$$\lim_{N \rightarrow \infty} [N^2 (\alpha\beta - \gamma^2)^{N-2} I_N] = \int_0^\infty \int_0^\infty e^{-u\beta - v\alpha} du dv = 1/\alpha\beta.$$

Now

$$E(r^2) = \frac{n}{4}(\alpha\beta - \gamma^2)^{\frac{n}{2}} I_{\frac{n}{2}+1} + \frac{n(n+2)}{4} \gamma^2 (\alpha\beta - \gamma^2)^{\frac{n}{2}} I_{\frac{n}{2}+3} \\ - \frac{n(\alpha\beta - \gamma^2)}{4\left(\frac{n}{2}+1\right)^2} \left[\left(\frac{n}{2}+1\right)^2 (\alpha\beta - \gamma^2)^{\frac{n}{2}-1} I_{\frac{n}{2}+1} \right] + \frac{n(n+2)\gamma^2}{4\left(\frac{n}{2}+2\right)^2} \left[\left(\frac{n}{2}+2\right)^2 (\alpha\beta - \gamma^2)^{\frac{n}{2}} I_{\frac{n}{2}+3} \right].$$

So, as n becomes indefinitely large, we get $E(r^2) \rightarrow \lambda^2/\alpha\beta$ to the first order of approximation.

REFERENCES TO LITERATURE

- AITKEN, A. C., and SILVERSTONE, H., 1941-42. "On the Estimation of Statistical Parameters", *Proc. Roy. Soc. Edin.*, A, LXI, 186-194.
- BHATTACHARYA, A., 1946. "On some Analogues of the Amount of Information and their Uses in Statistical Estimation", *Sankhya*, VIII, 1-14.
- , 1947. *Idem.*, *ibid.*, VIII, 201-218.
- , 1948. *Idem.*, *ibid.*, VIII, 315-328.
- FISHER, R. A., 1921. "Mathematical Foundations of Theoretical Statistics", *Phil. Trans. Roy. Soc.*, A, CCXXII, 309-368.

(Issued separately June 16, 1950)

VII.—Parallel Planes in a Riemannian V_n . By H. S. Ruse, The University, Leeds.

(MS. received June 9, 1949. Read November 7, 1949)

SYNOPSIS

In an earlier paper a description was given, in terms of classical projective geometry, of some of the properties of parallel fields of vector spaces (parallel planes) in a Riemannian V_n , and a detailed analysis was made of the case $n=4$. The present paper contains the corresponding formulæ for any n , though omits their projective interpretation. A parallel p -plane is said to be of nullity q when the p vectors of any normal basis contain q null and $p-q$ non-null vectors. The conditions of parallelism, namely that the covariant derivatives of the basis-vectors should depend linearly upon these vectors, are examined for any p and any $q(<p)$, and attention is thereafter mainly confined to the cases (i) n even, $q=\frac{1}{2}n-1$, $p=\frac{1}{2}n-1$ or $\frac{1}{2}n$; (ii) n odd, $q=\frac{1}{2}(n-3)$, $p=\frac{1}{2}(n-1)$, which possess exceptional features. In the former of these cases light is thrown upon the curious circumstance, noted in the previous paper, that the existence in a V_4 of a null parallel 1-plane necessitates the existence of parallel planes other than its conjugate. For a general n similar situations arise in the cases indicated.

THIS paper follows two others. One, by A. G. Walker,* initiated the theory of parallel fields of partially null vector spaces in a Riemannian V_n , and the other, by myself,† developed some of the geometrical and analytical consequences of his theory, with particular reference to the case $n=4$. The present paper gives the generalization to any n of results previously obtained for $n=4$.

The terminology is that of the previous papers. The V_n is assumed to be analytic and of fundamental tensor g_{ij} , the letters i, j, k, \dots being used for tensor suffixes (range 1, 2, . . . , n). The inner product $X_i Y^i \equiv g_{ij} X^i Y^j$ of any two vectors is denoted by (XY) or (YX) .

This paper is concerned solely with fields of vectors and of vector spaces, but as a rule the word *field* will be omitted. So, for example, a parallel field of vector spaces, or, as Walker calls them, a parallel field of planes, will be referred to as a *parallel plane*, or as a *parallel p -plane* when its dimensionality p needs explicit mention.

1. PARALLEL p -PLANE IN V_n

A necessary and sufficient condition (Walker, Theorem 3.3) that a set of p independent contravariant vectors $\lambda_a^i \equiv (\lambda_1^i, \dots, \lambda_p^i)$ should form a basis for a (contravariant) parallel p -plane V^p in V_n is that their covariant

* A. G. Walker, "On Parallel Fields of Partially Null Vector Spaces", *Quart. Journ. Math. (Oxford)*, XX, 1949, 135-145.

† H. S. Ruse, "On Parallel Fields of Planes in a Riemannian Space", *Quart. Journ. Math. (Oxford)*, XX, 1949, 218-234.

derivatives $\lambda_{\alpha,j}^i$ should be linear combinations of the vectors themselves, say

$$\lambda_{\alpha,j}^i \equiv A_{\alpha j}^{\beta} \lambda_{\beta}^i, \quad (1.1)$$

where the functions $A_{\alpha j}^{\beta}$ are covariant vectors of V_n , α, β being scalar indices. The identities (1.1) are called *recurrence-formulae* for the vectors λ_{α}^i .

If $\mu_{\alpha}^i \equiv t_{\alpha}^{\beta} \lambda_{\beta}^i$, where the scalars t_{α}^{β} are such that $\det |t_{\alpha}^{\beta}| \neq 0$, then the μ_{α}^i satisfy relations of the form (1.1) and constitute a new basis for the same parallel p -plane. The vectors of a basis may therefore be chosen, in an infinite number of ways, to be mutually orthogonal, and, when non-null, to be unit. Such a basis is called *normal*. For a given p -plane, the number q of null vectors is the same for every normal basis, and is called the *nullity* of the p -plane. It is easy to see that

$$q < \min(p, n-p), \quad (1.2)$$

and hence that $q < \frac{1}{2}n$, equality being possible only when n is even.

The *null part* of a parallel p -plane V^p is the vector space (or rather field of vector spaces), having as basis the q null vectors of any normal basis of V^p . It, too, is parallel (Walker, Theorem 3.1). When $q=0$, the null part of V^p is the 0 -plane, and contains only the zero vector.

The totality of vector fields which, at every point of V_n , are orthogonal to the vectors of V^p at that point, constitute the parallel $(n-p)$ -plane, *conjugate* to V^p (Walker, *ibid.*). It has the same null part as V^p .

2. THE PARALLEL n -PLANE: QUASI-NORMAL BASIS

The totality of contravariant vectors at any point of V_n constitute the vector n -space at that point, and the field of these spaces over V_n is the parallel n -plane V^n of V_n . By (1.2), the nullity of V^n is zero.

As a normal basis for the n -plane we may take any orthogonal ennuple. Let such an ennuple be $h_{\alpha}^i \equiv (h_1^i, \dots, h_n^i)$, where

$$(h_{\alpha} h_{\beta}) \equiv h_{\alpha}^i h_{\beta i} = c_{\alpha} \delta_{\alpha\beta} (c_{\alpha} = \pm 1) \quad (\text{not summed for } \alpha), \quad (2.1)$$

and let

$$h_i^{\alpha} \equiv c_{\alpha} h_{\alpha i} \quad (\text{not summed}). \quad (2.2)$$

Then by (2.1),

$$(h^{\alpha} h_{\beta}) = \delta_{\beta}^{\alpha}. \quad (2.3)$$

The corresponding recurrence-formulae are

$$h_{\alpha,j}^i = A_{\alpha j}^{\beta} h_{\beta}^i.$$

In this case the coefficients A_{ij}^b are given by

$$A_{ij}^b = \gamma_{a..i}^b h_j^a,$$

or by

$$A_{ij}^b h_i^j = \gamma_{a..i}^b,$$

where

$$\gamma_{a..i}^b \equiv h_{a..j}^i h_i^b h_j^a = e_b \gamma_{ab..i},$$

$\gamma_{ab..i}$ being Ricci's coefficients of rotation.

If instead of taking the orthogonal ennuple h_a^i we take a *quasi-orthogonal ennuple** consisting partly of null and partly of non-null vectors, we obtain what may be called a *quasi-normal* basis for V^n . This may be done as follows.

Let r be any positive integer $< \frac{1}{2}n$, and choose from among the h_a^i any two non-overlapping sets of r of them. Without loss of generality these may be taken to be the first r and the last r , so that the vectors are divided into three consecutive sets. Writing $\dagger R, S, R'$ for the three ranges of numbers indicated in the diagram

$$\left. \begin{array}{ccc} (1, 2, \dots, r) & (r+1, r+2, \dots, n-r) & (n-r+1, \dots, n) \\ R & S & R' \end{array} \right\}, \quad (2.4)$$

we then have

$$\begin{aligned} g_{ij} &= \sum_{a=1}^n e_a h_{ai} h_{aj} \\ &= \sum_{a \in R} e_a h_{ai} h_{aj} + \sum_{a \in S} e_a h_{ai} h_{aj} + \sum_{a \in R'} e_a h_{ai} h_{aj}. \end{aligned} \quad (2.5)$$

Thus R, S, R' are respectively the first r , the middle $n-2r$, and the last r numbers of the set $1, 2, \dots, n$.

For any given a , let $a' = a + n - r$. Then when $a \in R$, clearly $a' \in R'$, in conformity with (2.5). Let

$$\left. \begin{aligned} \xi_a^i &\equiv \frac{1}{\sqrt{2}} (e_a^i h_a^i + i e_a^i h_{a'}^i) \\ \zeta_a^i &\equiv \frac{1}{\sqrt{2}} (e_a^i h_a^i - i e_a^i h_{a'}^i) \end{aligned} \right\} \quad (a \in R) \quad (2.6)$$

(not summed for a, a'), where $e_a^i = 1$ or $i[\equiv \sqrt{(-1)}]$ according as $e_a = +1$ or -1 . Then the $2r$ vectors $h_a^i, h_{a'}^i$ are replaced by the $2r$ vectors ξ_a^i, ζ_a^i . The latter, namely $\xi_a^i, \zeta_a^i, \dots, \xi_{a'}^i, \zeta_{a'}^i$, could alternatively have been denoted by $\zeta_{a'}^i$ (or, for that matter, by $\xi_{a'}^i$), with a' running over R' , but

* Y. C. Wong, "Quasi-orthogonal Ennuple of Congruences in a Riemannian Space", *Ann. of Math.*, XLVI, 1945, 158-173.

† The idea of representing suffix-ranges by single capital letters is due to A. G. Walker.

the notation adopted, which is similar to that used by Wong, proves to be more convenient.

It is obvious that the vectors $\xi_a^i, \zeta^{\alpha i}$, may or may not be real. They are real only when each $e_a = +1$ and each $e_\alpha = -1$. For a Riemannian space of positive-definite metric they will certainly all be unreal

From (2.6), (2.5) and (2.2) we get

$$g_{ij} = \xi_{\alpha i} \zeta_j^\alpha + \zeta_i^\alpha \xi_{\alpha j} + h_i^\sigma h_{\sigma j} \quad (\alpha \in R, \sigma \in S), \quad (2.7)$$

where the summation convention now applies to Greek indices, each taken over its appropriate range. Also from (2.6), (2.1), (2.2) and (2.3) it follows that *

$$(\xi_\alpha \xi_\beta) = 0 = (\zeta^\alpha \zeta^\beta) \quad [(a, \beta) \in R], \quad (2.8)$$

$$(\xi_\alpha \zeta^\beta) = \delta_\alpha^\beta, \quad (2.9)$$

$$(\xi_\alpha h_\sigma) = 0 = (\zeta^\alpha h_\sigma) \quad (\sigma \in S), \quad (2.10)$$

$$(h_\sigma h^\tau) = \delta_\sigma^\tau \quad [(\sigma, \tau) \in S]. \quad (2.11)$$

Thus, by (2.8), the ξ 's and ζ 's are all null vectors, those of each set being mutually orthogonal. Also, by (2.9), each ξ_α^i is orthogonal to all but one of the ζ 's, namely the one having the same index α . By (2.10), the ξ 's and ζ 's are all orthogonal to all the unit vectors h_σ^i . Equation (2.11) is simply (2.3) applied to the "middle" set of the vectors h_σ^i , and expresses the fact that they are all unit and mutually orthogonal.

The $2r$ null vectors $\xi_\alpha^i, \zeta^{\alpha i}$, and the $(n-2r)$ unit vectors h_σ^i , form a *quasi-orthogonal ennuple of nullity*† r , the vectors all being mutually orthogonal except each ξ_α^i and the corresponding $\zeta^{\alpha i}$. If n is even and r is taken equal to $\frac{1}{2}n$, all the vectors are null.

With these vectors as a quasi-normal basis for V^n , the recurrence-formulae take the form

$$\xi_{\alpha, j}^i = A_{\alpha j}^\beta \xi_\beta^i + B_{\alpha \beta j} \zeta^{\beta i} + F_{\alpha \tau j} h^{\tau i},$$

$$\zeta_{, j}^{\alpha i} = C_{j \beta}^{\alpha} \xi_\beta^i + D_{\beta j}^{\alpha} \zeta^{\beta i} + G_{\tau j}^{\alpha} h^{\tau i},$$

$$h_{\sigma, j}^i = P_{\sigma j}^\beta \xi_\beta^i + Q_{\sigma \beta j} \zeta^{\beta i} + H_{\sigma \tau j} h^{\tau i},$$

where $(\alpha, \beta) \in R$ and $(\sigma, \tau) \in S$, the Greek indices in the coefficients $A_{\beta j}^\alpha, \dots, H_{\sigma \tau j}$ being written in the covariant or contravariant positions in such a way as to conform with the Greek indices of the ξ 's, ζ 's and h 's.

From (2.8), \dots , (2.11) we quickly find that $B_{\alpha \beta j}$ and $C_{j \beta}^\alpha$ are both

* The notation $(\alpha, \beta) \in R$ is an abbreviation for " $\alpha \in R$ and $\beta \in R$ ".

† Wong, *loc. cit.* This use of *nullity* is different from that of § 1, where it was used in connection with parallel planes. The latter meaning is the one to be understood in this paper, except in the present instance, which is the only occasion upon which the word is used in Wong's sense.

skew in α, β ; that $D_{\beta\gamma}^{\alpha} = -A_{\beta\gamma}^{\alpha}$; that $P_{\sigma\gamma}^{\alpha} = -G_{\sigma\gamma}^{\alpha}$; that $Q_{\sigma\beta\gamma} = -F_{\beta\sigma\gamma}$; and that $H_{\sigma\tau\gamma}$ is skew in σ, τ . Thus the recurrence-formulae for the parallel n -plane, which are in fact identities satisfied by the vectors of the given quasi-orthogonal ennuple, reduce to *

$$\left. \begin{aligned} (a) \quad \xi_{\alpha,j}^i &= A_{\alpha\beta}^{\beta} \xi_{\beta}^i + B_{\alpha\beta\gamma} \zeta_{\beta}^{\alpha i} + F_{\sigma\tau\gamma} h^{\tau i}, \\ (b) \quad \zeta_{\alpha,j}^{\alpha i} &= C_{\beta\gamma}^{\alpha\beta} \xi_{\beta}^i - A_{\beta\gamma}^{\alpha} \gamma^{\beta i} + G_{\sigma\tau\gamma} h^{\tau i}, \\ (c) \quad h_{\sigma,j}^i &= -G_{\sigma\gamma}^{\beta} \xi_{\beta}^i - F_{\beta\sigma\gamma} \zeta_{\beta}^{\alpha i} + H_{\sigma\tau\gamma} h^{\tau i}, \end{aligned} \right\} \quad (2.12)$$

where

$$B_{\alpha\beta\gamma} = -B_{\beta\alpha\gamma}, \quad C_{\beta\gamma}^{\alpha\beta} = -C_{\gamma\beta}^{\alpha\alpha}, \quad H_{\sigma\tau\gamma} = -H_{\tau\sigma\gamma}, \quad (2.13)$$

and

$$(\alpha, \beta) \in R, \quad (\sigma, \tau) \in S.$$

Equations (2.12) will be referred to as the *general identities* for the quasi-orthogonal ennuple $(\xi_{\alpha}^i, \zeta_{\alpha}^{\alpha i}, h_{\sigma}^i)$.

When $n=4$ and $r=2$, these give the formulae (4.12) of my previous paper. There the ξ_{α}^i are denoted by ξ^i, η^i , and the $\zeta_{\alpha}^{\alpha i}$ by ξ'^i, η'^i , all four vectors of the quasi-orthogonal ennuple being null. The coefficients here denoted by $A_{\alpha\beta}^{\beta}$ are there represented by A_j, B_j, C_j, D_j , while $B_{12\gamma}$ and C_j^{12} are represented by $-W_j'$ and $-W_j$.

3. PARALLEL $(r+s)$ -PLANE OF NULLITY r

Any normal basis for a parallel $(r+s)$ -plane of nullity r contains r null and s unit vectors. By (1.2), $r < \min(r+s, n-r-s)$, and so, in particular, $r < n-r-s$, whence $r+s < n-r$. Let R, S_1, S_2 be the suffix-ranges defined by

$$\left. \begin{aligned} & (1, 2, \dots, r) \quad (r+1, r+2, \dots, r+s) \quad (r+s+1, \dots, n-r) \\ & \quad R \quad \quad \quad S_1 \quad \quad \quad S_2 \end{aligned} \right\}, \quad (3.1)$$

S_2 being non-existent when it happens that $r+s=n-r$. S_1 and S_2 together comprise the range hitherto denoted by S , so $S=(S_1, S_2)$.

Let the null vectors of a given normal basis for the $(r+s)$ -plane be $\xi_{\alpha}^i \equiv (\xi_1^i, \dots, \xi_r^i)$ (so $\alpha \in R$), and the non-null be $h_{\mu}^i \equiv (h_{r+1}^i, \dots, h_{r+s}^i)$ (so $\mu \in S_1$). Then the recurrence-formulae for the parallel $(r+s)$ -plane have the form

$$\xi_{\alpha,j}^i = E_{\alpha\beta}^{\beta} \xi_{\beta}^i + K_{\sigma\tau\gamma} h^{\tau i}, \quad (3.2)$$

$$h_{\mu,j}^i = M_{\alpha\beta}^{\beta} \xi_{\beta}^i + N_{\sigma\tau\gamma} h^{\tau i}, \quad (3.3)$$

where $(\alpha, \beta) \in R$ and $(\mu, \nu) \in S_1$, these merely stating that the covariant derivatives of the basis-vectors are linearly expressible in terms of those vectors.

* Cf. Wong, *loc. cit.*, equation (3.2).

The orthogonal vectors ξ_α^i, h_μ^i may be taken as $r+s$ of the vectors of a quasi-orthogonal ennuple, which may be completed by the choice of r null vectors ζ^α and of $n-2r-s$ unit vectors, say $k_\theta^i \equiv (k_{r+s+1}^i, \dots, k_{n-r}^i)$ (so that $\theta \in S_2$). The last set might be denoted by h_θ^i because they, with the h_μ^i , together make up the whole set of $n-2r$ unit vectors h_σ^i [$\sigma \in (S_1, S_2)$] of the quasi-orthogonal ennuple, but they are denoted by k_θ^i to emphasize the fact that they are not part of the basis of the given $(r+s)$ -plane. We thus have the sets of vectors $\xi_\alpha^i, \zeta^\alpha, h_\mu^i, k_\theta^i$, respectively $r, r, s, n-2r-s$ in number, of which ξ_α^i, h_μ^i define the parallel $(r+s)$ -plane, the ζ 's and k 's being auxiliary. The general identities (2.12) for the whole ennuple are now

$$\left. \begin{aligned} (a) \quad \xi_{\alpha,j}^i &= A_{\alpha j}^\beta \xi_\beta^i + B_{\alpha \beta j} \zeta^{\beta i} + F_{\alpha \mu j} h_\mu^i + F_{\alpha \theta j} k_\theta^i, \\ (b) \quad \zeta_{,j}^\alpha &= C_{\alpha j}^\beta \xi_\beta^i - A_{\beta j}^\alpha \zeta^{\beta i} + G_{\mu j}^\alpha h_\mu^i + G_{\theta j}^\alpha k_\theta^i, \\ (c) \quad h_{\mu,j}^i &= -G_{\mu j}^\beta \xi_\beta^i - F_{\beta \mu j} \zeta^{\beta i} + H_{\mu \nu j} h_\nu^i + H_{\mu \theta j} k_\theta^i, \\ (d) \quad k_{\theta,j}^i &= -G_{\theta j}^\beta \xi_\beta^i - F_{\beta \theta j} \zeta^{\beta i} + H_{\theta \nu j} h_\nu^i + H_{\theta \phi j} k_\phi^i, \end{aligned} \right\} \quad (3.4)$$

where $(\alpha, \beta) \in R, (\mu, \nu) \in S_1, (\theta, \phi) \in S_2$. These equations are the same as (2.12), but with the range S of the indices σ, τ in (2.12) split into the two ranges S_1, S_2 , the h_σ^i , for $\sigma \in S_2$, being written as k 's.

For the ξ 's and h 's to define a parallel $(r+s)$ -plane, (3.4) (a) and (c) must be respectively of the form (3.2), (3.3). Hence, in (3.4) (a), both $B_{\alpha \beta j}$ and $F_{\alpha \theta j}$ are zero, while in (3.4) (c) both $F_{\beta \mu j}$ and $H_{\mu \theta j}$ are zero. Thus all the F 's are zero, and the right-hand side of (3.4) (a) reduces to the ξ -term. Also, because $H_{\sigma \tau j}$ is skew in σ, τ , by (2.13), $H_{\theta \nu j}$ in (3.4) (d) is zero because $H_{\mu \theta j}$ is so. Hence recurrence-formulae appropriate to the case of a parallel $(r+s)$ -plane of nullity r are of the form

$$\left. \begin{aligned} (a) \quad \xi_{\alpha,j}^i &= A_{\alpha j}^\beta \xi_\beta^i, \\ (b) \quad \zeta_{,j}^\alpha &= C_{\alpha j}^\beta \xi_\beta^i - A_{\beta j}^\alpha \zeta^{\beta i} + G_{\mu j}^\alpha h_\mu^i + G_{\theta j}^\alpha k_\theta^i, \\ (c) \quad h_{\mu,j}^i &= -G_{\mu j}^\beta \xi_\beta^i + H_{\mu \nu j} h_\nu^i, \\ (d) \quad k_{\theta,j}^i &= -G_{\theta j}^\beta \xi_\beta^i + H_{\theta \nu j} k_\nu^i, \end{aligned} \right\} \quad (3.5)$$

where $(\alpha, \beta) \in R, (\mu, \nu) \in S_1, (\theta, \phi) \in S_2$ and $k_\theta^i = e_\theta k_\theta^i$.

Equations (3.5) (a), (c) are the ones defining the given $(r+s)$ -plane Π . The fact that (3.5) (a) involves only the ξ_α^i shows analytically that the null part ω of the parallel $(r+s)$ -plane is also parallel, as stated in § 1 above. The non-null part, defined by the h_μ^i , is in general not parallel, because the right-hand side of (3.5) (c) involves the ξ 's as well as the h 's: indeed, "non-null part" is really a misnomer, because a partially null parallel plane in general has no uniquely defined non-null part, the h_μ^i being replaceable by linear combinations of themselves and of the ξ_α^i . Equations

(3.5) (a), (d) show that the r null vectors ξ_a^i and the $n-2r-s$ unit vectors k_θ^i , all of which are orthogonal to one another, are a normal basis of a parallel $(n-r-s)$ -plane. This is the $(n-r-s)$ -plane P conjugate to the given $(r+s)$ -plane Π , its null part, which has the ξ_a^i for basis, being the same as the null part ω of Π . Similarly, (3.5) (a), (c), (d) together show that the $(n-r)$ -plane p of basis $(\xi_a^i, k_\mu^i, k_\theta^i)$ is parallel. It is the plane conjugate to the null part ω of the original plane. All of the planes intersect in ω and are contained in p .

Equations (3.5) (b) are auxiliary, the four sets of equations (3.5) being recurrence-formulae for the whole n -plane V^n , which, of course, contains Π , P , ω and p .

Equations (3.5) are the most general ones for a partially null parallel plane in V_n . Various special cases are obtainable from them. For example, if the given parallel plane is wholly null (case $s=0$), then there are no vectors k_μ^i but there are $n-2r$ vectors k_θ^i . In (3.5) (b) the term in k_μ^i disappears, and (3.5) (c) is non-existent. Otherwise the equations remain formally the same.* Again, if it happens that n is even and $r=\frac{1}{2}n$, all the vectors of the quasi-orthogonal ennuple are null, and the general identities reduce simply to

$$\left. \begin{aligned} (a) \quad \xi_{a,j}^i &= A_{aj}^\beta \xi_\beta^i, \\ (b) \quad \zeta_{,j}^{\alpha i} &= C_{,j}^{\alpha\beta} \xi_\beta^i - A_{\beta j}^\alpha \zeta^{\beta i}, \end{aligned} \right\} \quad (3.6)$$

with α, β running from 1 to $\frac{1}{2}n$. Equations (3.6) (a) are the recurrence-formulae for the null parallel $\frac{1}{2}n$ -plane, which is self-conjugate, (3.6) (b) being auxiliary. If it also happens that $C^{\alpha\beta} \equiv 0$, then the $\zeta^{\alpha i}$ are a basis of a second null parallel $\frac{1}{2}n$ -plane having only the 0-plane in common with the former.

The case when n is even and $r=\frac{1}{2}n$ thus contains nothing of exceptional interest. The case when n is even and $r=\frac{1}{2}n-1$ has, however, certain peculiar features, and the same is true when n is odd and $r=\frac{1}{2}(n-3)$. These are considered in the following sections.

4. NULL PARALLEL $(\frac{1}{2}n-1)$ -PLANE IN V_n (n EVEN)

Suppose n even, and let

$$\begin{aligned} M &\equiv 1, 2, 3, \dots, (\tfrac{1}{2}n-1); \\ (M, \tfrac{1}{2}n) &\equiv 1, 2, 3, \dots, (\tfrac{1}{2}n-1), \tfrac{1}{2}n. \end{aligned}$$

Suppose also that V_n admits a null parallel $(\frac{1}{2}n-1)$ -plane having as normal

* The vectors k_θ^i are auxiliary, and pairs of them could be replaced by pairs of auxiliary null vectors defined in terms of them by formulae similar to (2.6), in which case (3.5) would be modified in form.

basis a set of independent orthogonal null vectors $\xi_\rho^i \equiv [\xi_1^i, \dots, \xi_{(\frac{1}{2}n-1)}^i]$. The recurrence-formulae for this plane are then of the form

$$\xi_{\rho,j}^i = A_{\rho j}^\sigma \xi_\sigma^i \quad [(\rho, \sigma) \in M]. \quad (4.1)$$

Form a quasi-orthogonal ennuple, consisting entirely of null vectors, by taking, with the ξ_ρ^i , a further null vector $\xi_{(\frac{1}{2}n)}^i$ —thereby making a set of $\frac{1}{2}n$ ξ -vectors $\xi_\alpha^i \equiv [\xi_\rho^i, \xi_{(\frac{1}{2}n)}^i]$ —and $\frac{1}{2}n$ null vectors $\zeta^{\alpha i}$. By (2.12), the general identities for this ennuple are of the form

$$\xi_{\alpha,j}^i = A_{\alpha j}^\beta \xi_\beta^i + B_{\alpha\beta j} \zeta^{\beta i}, \quad (4.2)$$

$$\zeta^{\alpha i}_j = C^{\alpha\beta}_j \xi_\beta^i - A_{\beta j}^\alpha \zeta^{\beta i}, \quad (4.3)$$

where $(\alpha, \beta) \in (M, \frac{1}{2}n)$, $B_{\alpha\beta j}$ and $C^{\alpha\beta}_j$ being skew in α, β .

Equations (4.1) must be the same as the first $\frac{1}{2}n - 1$ of equations (4.2), namely

$$\xi_{\rho,j}^i = A_{\rho j}^\beta \xi_\beta^i + B_{\rho\beta j} \zeta^{\beta i} \quad \left[\begin{array}{l} \rho \in M \\ \beta \in (M, \frac{1}{2}n) \end{array} \right], \quad (4.4)$$

the $(\frac{1}{2}n)$ th equation (4.2) being

$$\xi_{(\frac{1}{2}n),j}^i = A_{(\frac{1}{2}n),j}^\beta \xi_\beta^i + B_{(\frac{1}{2}n)\beta j} \zeta^{\beta i}. \quad (4.5)$$

Comparison of (4.4) with (4.1) gives

$$A_{\rho j}^{(\frac{1}{2}n)} = 0 \quad (\rho \in M) \quad (4.6)$$

and

$$B_{\rho\beta j} = 0 \quad [\rho \in M, \beta \in (M, \frac{1}{2}n)]. \quad (4.7)$$

The last equation shows that, for any fixed j , the first $\frac{1}{2}n - 1$ rows of the $(\frac{1}{2}n) \times (\frac{1}{2}n)$ matrix $B_{\alpha\beta j}$ are zero. Therefore, since the matrix is skew, it is wholly zero, so

$$B_{\alpha\beta j} = 0. \quad (4.8)$$

Thus (4.4) and (4.5) become

$$\xi_{\rho,j}^i = A_{\rho j}^\sigma \xi_\sigma^i \quad [(\rho, \sigma) \in M], \quad (4.9)$$

$$\xi_{(\frac{1}{2}n),j}^i = A_{(\frac{1}{2}n),j}^\beta \xi_\beta^i \quad [\beta \in (M, \frac{1}{2}n)], \quad (4.10)$$

(4.9) being the same as (4.1). These two equations together show that the vectors $\xi_\alpha^i \equiv [\xi_\rho^i, \xi_{(\frac{1}{2}n)}^i]$ are a (normal) basis of a (null) parallel $\frac{1}{2}n$ -plane.

Now consider the equation obtained by taking $\alpha = \frac{1}{2}n$ in (4.3), namely

$$\zeta_{j}^{(\frac{1}{2}n)i} = C^{(\frac{1}{2}n)\beta}_j \xi_\beta^i - A_{\beta j}^{(\frac{1}{2}n)} \zeta^{\beta i} \quad [\beta \in (M, \frac{1}{2}n)]. \quad (4.11)$$

By (4.6), the last term on the right of (4.11) is equal to $A_{(\frac{1}{2}n),j}^{(\frac{1}{2}n)} \zeta^{(\frac{1}{2}n)i}$, and so involves only the vector $\zeta^{(\frac{1}{2}n)i}$ out of the set $\zeta^{\alpha i}$. Also, since $C^{\alpha\beta}_j$ is skew in α, β , we have $C^{(\frac{1}{2}n)(\frac{1}{2}n)}_j = 0$, so the first term on the right is equal to $C^{(\frac{1}{2}n)\rho}_j \xi_\rho^i$, with ρ summing over M only. Thus (4.11) is

$$\zeta_{j}^{(\frac{1}{2}n)i} = C^{(\frac{1}{2}n)\rho}_j \xi_\rho^i - A_{(\frac{1}{2}n),j}^{(\frac{1}{2}n)} \zeta^{(\frac{1}{2}n)i}, \quad (4.12)$$

and this, with (4.9), shows that ξ_ρ^i and $\zeta^{(i)n}$ are a (normal) basis of another (null) parallel $\frac{1}{2}n$ -plane.

We thus have in V_n the given null $(\frac{1}{2}n-1)$ -plane having $\xi_\rho^i (\rho \in M)$ as basis, and the two null $\frac{1}{2}n$ -planes having $[\xi_\rho^i, \xi_{(\frac{1}{2}n)}^i]$ and $[\xi_\rho^i, \zeta^{(i)n}]$ as respective bases. The $\frac{1}{2}n$ -planes intersect in the given $(\frac{1}{2}n-1)$ -plane, and all are contained in the parallel $(\frac{1}{2}n+1)$ -plane having $[\xi_\rho^i, \xi_{(\frac{1}{2}n)}^i, \zeta^{(i)n}]$ as basis, which is non-normal because, by (2.9), $\xi_{(\frac{1}{2}n)}^i$ and $\zeta^{(i)n}$ are not orthogonal. This $(\frac{1}{2}n+1)$ -plane is the parallel plane conjugate to the given $(\frac{1}{2}n-1)$ -plane, and is of nullity $\frac{1}{2}n-1$.

So we have the theorem: *When n is even and V_n admits a null parallel $(\frac{1}{2}n-1)$ -plane, it also admits two null parallel $\frac{1}{2}n$ -planes intersecting in the given $(\frac{1}{2}n-1)$ -plane. These planes are all contained in the parallel $(\frac{1}{2}n+1)$ -plane of nullity $(\frac{1}{2}n-1)$ conjugate to the given $(\frac{1}{2}n-1)$ -plane.*

The special case of this theorem for $n=4$ was given in § 6 of my previous paper.

5. PARALLEL PLANES OF NULLITY $(\frac{1}{2}n-1)$ IN V_n (n EVEN)

If, in V_n (n even), we have a parallel $(r+s)$ -plane of nullity r , where $r = \frac{1}{2}n-1$, then, because $r < n-r-s$, by (1.2), we have $s < 2$. Hence if a parallel plane is of nullity $(\frac{1}{2}n-1)$, it is either (i) a $(\frac{1}{2}n+1)$ -plane; or (ii) a $\frac{1}{2}n$ -plane; or (iii) a $(\frac{1}{2}n-1)$ -plane. Case (iii) was dealt with in § 4 above, and so, in effect, was case (i), since the conjugate of a parallel $(\frac{1}{2}n+1)$ -plane of nullity $(\frac{1}{2}n-1)$ is a null $(\frac{1}{2}n-1)$ -plane. The only distinct case that remains for consideration is therefore (ii).

Suppose, then, that V_n admits a $\frac{1}{2}n$ -plane Π of nullity $(\frac{1}{2}n-1)$. Take a normal basis for Π consisting of the null vectors $\xi_\rho^i \equiv [\xi_1^i, \dots, \xi_{(\frac{1}{2}n)}^i]$ and of the one unit vector h^i (indicator $e_1 = \pm 1$). Form a quasi-orthogonal ennuple by taking, with ξ_ρ^i and h^i , a set of $\frac{1}{2}n-1$ null vectors ζ^{ai} and another unit vector, k^i say (indicator e_2). Then by (2.12) the general identities for this ennuple are of the form

$$\xi_{\rho,j}^i = A_{\rho j}^\sigma \xi_\sigma^i + B_{\rho j}^\sigma \zeta^{a\sigma} + F_{\rho 1j} e_1 h^i + F_{\rho 2j} e_2 k^i, \quad (5.1)$$

$$\zeta_{\rho,j}^{ai} = C_{\rho j}^\sigma \xi_\sigma^i - A_{\rho j}^\sigma \zeta^{a\sigma} + G_{\rho 1j} e_1 h^i + G_{\rho 2j} e_2 k^i, \quad (5.2)$$

$$h_{,j}^i = -G_{1j}^\sigma \xi_\sigma^i - F_{\sigma 1j} \zeta^{a\sigma} + H_{12j} e_2 k^i, \quad (5.3)$$

$$k_{,j}^i = -G_{2j}^\sigma \xi_\sigma^i - F_{\sigma 2j} \zeta^{a\sigma} + H_{21j} e_1 h^i, \quad (5.4)$$

where ρ, σ run from 1 to $\frac{1}{2}n-1$, and where $H_{21j} = -H_{12j}$.

Inasmuch as (ξ_ρ^i, h^i) are a basis of a parallel plane, their covariant derivatives are linear functions of them alone, so we have, by (5.1),

$$B_{\rho\sigma j} = 0 = F_{\rho 2j},$$

and, by (5.3),

$$F_{\sigma 1j} = 0 = H_{12j}.$$

Thus (5.1), . . . , (5.4) reduce to

$$\xi_{n,j}^i = A_{nj}^\sigma \xi_\sigma^i, \quad (5.5)$$

$$\zeta_{n,j}^i = C_{nj}^\sigma \xi_\sigma^i - A_{nj}^\sigma \zeta_\sigma^i + G_{nj}^\sigma e_1 h^i + G_{nj}^\sigma e_2 h^i, \quad (5.6)$$

$$h_{n,j}^i = -G_{nj}^\sigma \xi_\sigma^i, \quad (5.7)$$

$$k_{n,j}^i = -G_{nj}^\sigma \xi_\sigma^i. \quad (5.8)$$

Of these, (5.6) is auxiliary and requires no further consideration. Equations (5.5), (5.7) are the recurrence-formulae for the given $\frac{1}{2}n$ -plane Π , which is based on ξ_σ^i and h^i . Equations (5.5), (5.8) show that there is another parallel $\frac{1}{2}n$ -plane, based on ξ_σ^i and k^i , intersecting Π in its null part, *i.e.* in the null $(\frac{1}{2}n - 1)$ -plane ω based on ξ_σ^i . By (5.5), or by Walker's Theorem 3.1, this null part is by itself parallel, and therefore, by § 4, V_n also admits two null parallel $\frac{1}{2}n$ -planes intersecting in ω . Moreover, if a, b are any constants, we have, by (5.7) and (5.8),

$$(ah^i + bk^i),_j = -(aG_{1j}^\sigma + bG_{2j}^\sigma)\xi_\sigma^i,$$

and this, with (5.5), shows that the vectors $\xi_\sigma^i, ah^i + bk^i$, for any constant a, b , are a basis of a parallel $\frac{1}{2}n$ -plane. If $ah^i + bk^i$ is non-null, as will in general be the case, the $\frac{1}{2}n$ -plane is of nullity $(\frac{1}{2}n - 1)$; but if the ratio a/b has either of the values that make $ah^i + bk^i$ null, then the $\frac{1}{2}n$ -plane is wholly null. These two values of a/b in fact give the two null $\frac{1}{2}n$ -planes intersecting in ω already referred to. All of the planes are contained in the parallel $(\frac{1}{2}n + 1)$ -plane of nullity $(\frac{1}{2}n - 1)$ conjugate to ω .

Summarizing these results, we have: *If n is even and the V_n admits a parallel $\frac{1}{2}n$ -plane of nullity $(\frac{1}{2}n - 1)$, then (i) its null part ω is a null parallel $(\frac{1}{2}n - 1)$ -plane; (ii) it admits a pencil of parallel $\frac{1}{2}n$ -planes all intersecting in ω ; of these, all except two are of nullity $(\frac{1}{2}n - 1)$, the exceptional two being wholly null $\frac{1}{2}n$ -planes; (iii) all the planes are contained in the parallel $(\frac{1}{2}n + 1)$ -plane conjugate to ω , which is of nullity $(\frac{1}{2}n - 1)$.*

The special case of this theorem for $n = 4$ was given in § 7 of my previous paper.

6. PARALLEL PLANES OF NULLITY $\frac{1}{2}(n - 3)$ IN V_n (n ODD)

When n is even, the greatest possible nullity of a plane is $\frac{1}{2}n$. It was seen above that a parallel plane possessing this maximum nullity had no other properties of special interest, but that parallel planes of nullity $(\frac{1}{2}n - 1)$, one less than the maximum, had exceptional properties.

When n is odd, the maximum nullity is $\frac{1}{2}(n - 1)$. Here again it turns out that planes having this nullity are not of special note, but that exceptional features do arise when the nullity is one less than this, namely $\frac{1}{2}(n - 3)$.

For an $(r+s)$ -plane of nullity r we have $r < n-r-s$, and hence, if $r = \frac{1}{2}(n-3)$, we have $s < 3$. The only possibilities are therefore:

- (i) $r+s = \frac{1}{2}(n-3)$, $s=0$; (ii) $r+s = \frac{1}{2}(n-1)$, $s=1$;
 (iii) $r+s = \frac{1}{2}(n+1)$, $s=2$; (iv) $r+s = \frac{1}{2}(n+3)$, $s=3$.

As the conjugate of a null $\frac{1}{2}(n-3)$ -plane is a $\frac{1}{2}(n+3)$ -plane of nullity $\frac{1}{2}(n-3)$, possibility (iv) is equivalent to possibility (i). Similarly (iii) is equivalent to (ii). Therefore the only distinct cases that need consideration are (i) and (ii), and it will be found that only (ii) has exceptional features.

We first obtain general identities in the form needed in the present section. Let

$$\left. \begin{aligned} N &\equiv 1, 2, 3, \dots, \frac{1}{2}(n-3), \\ (N, m) &\equiv 1, 2, 3, \dots, \frac{1}{2}(n-3), \frac{1}{2}(n-1) = m. \end{aligned} \right\} \quad (6.1)$$

In V_n choose a quasi-orthogonal ennuple consisting of $m [= \frac{1}{2}(n-1)]$ null vectors ξ_a^i , m other null vectors $\zeta^{\alpha i}$, and one non-null vector h^i (indicator e). Then the general identities (2.12) have the form

$$\left. \begin{aligned} (a) \quad \xi_{a,j}^i &= A_{aj}^\beta \xi_\beta^i + B_{aj\beta} \zeta^{\beta i} + F_{aj} e h^i, \\ (b) \quad \zeta_{\alpha,j}^i &= C_{\alpha j}^\beta \xi_\beta^i - A_{\alpha j}^\beta \zeta^{\beta i} + G_j^\alpha e h^i, \\ (c) \quad h_{,j}^i &= -G_j^\beta \xi_\beta^i - F_{\beta j} \zeta^{\beta i}, \end{aligned} \right\} \quad (6.2)$$

where $(\alpha, \beta) \in (N, m)$, the coefficients $F_{\alpha j}$, G_j^α of (2.12) being here written F_{aj} , G_j^α because there is only the one non-null vector h^i . There is no term in (6.2) (c) corresponding to h^i because $H_{\sigma j}$ in (2.12) (c) is skew in σ, τ .

Suppose now that we have in V_n a null parallel $\frac{1}{2}(n-3)$ -plane (case (i) above), having as basis the null vectors $\xi_\rho^i (\rho \in N)$. Then the recurrence-formulae for the plane are of the form

$$\boxed{\xi_{\rho,j}^i = A_{\rho j}^\sigma \xi_\sigma^i} \quad [(\rho, \sigma) \in N]. \quad (6.3)$$

Form a quasi-orthogonal ennuple by taking, with the given vectors ξ_ρ^i , a vector ξ_m^i [$m = \frac{1}{2}(n-1)$], thereby making a set $\xi_a^i \equiv (\xi_\rho^i, \xi_m^i)$, and also $\frac{1}{2}(n-1)$ null vectors $\zeta^{\alpha i}$ and one unit vector h^i . The vectors of this ennuple will satisfy identities of the form (6.2), and (6.2) (a) must therefore include (6.3). Rewriting (6.2) (a) for the cases $\alpha \neq \frac{1}{2}(n-1)$ and $\alpha = \frac{1}{2}(n-1) = m$, we get

$$\xi_{\rho,j}^i = A_{\rho j}^\beta \xi_\beta^i + B_{\rho j\beta} \zeta^{\beta i} + F_{\rho j} e h^i, \quad (6.4)$$

$$\xi_{m,j}^i = A_{mj}^\beta \xi_\beta^i + B_{mj\beta} \zeta^{\beta i} + F_{mj} e h^i. \quad (6.5)$$

Equation (6.4) must be the same as (6.3), and so

$$B_{\rho\beta} = 0 = F_{\rho\beta} \quad [\rho \in N, \beta \in (N, m)].$$

Because $B_{\alpha\beta}$ is skew in α, β , it follows that

$$B_{\alpha\beta} = 0.$$

Thus (6.4) becomes (6.3), (6.5) reduces to

$$\boxed{\xi_{m,j}^i = A_m^\beta \xi_\beta^i + F_{mj} e h^i}, \quad (6.6)$$

(6.2) (b) remains unaltered, and (6.2) (c) becomes

$$\boxed{h^i_{,j} = -G_j^\beta \xi_\beta^i - F_{mj} \zeta^{mi}}, \quad (6.7)$$

the last equation involving only the one ζ -vector ζ^{mi} .

The recurrence-formulae for the case at present under consideration are the three "boxed" equations (6.3), (6.6), (6.7), with (6.2) (b) as an auxiliary.

Equation (6.3) is the original recurrence-relation for the given null $\frac{1}{2}(n-3)$ -plane. The three boxed equations show that the vectors ξ_ρ^i , ξ_m^i , ζ^{mi} and h^i are a (non-normal) basis for a parallel $\frac{1}{2}(n+3)$ -plane. It is the $\frac{1}{2}(n+3)$ -plane conjugate to the given null $\frac{1}{2}(n-3)$ -plane.

In general the only other parallel planes in V_n are the o-plane and n -plane, so this case, unlike the corresponding one for even n , is not exceptional. Exceptional features might arise by accident, so to speak; for example if F_{mj} happened to be identically zero. In that case, by (6.3) and (6.6), the V_n would also admit a null parallel $\frac{1}{2}(n-1)$ -plane based on the vectors $\xi_\alpha^i \equiv (\xi_\rho^i, \xi_m^i)$, and, by (6.3) and (6.7), a parallel $\frac{1}{2}(n-1)$ -plane of nullity $\frac{1}{2}(n-3)$ based on (ξ_ρ^i, h^i) .

This brings us to the last case to be considered in the present paper. Suppose, without reference to the remark just made, that a V_n (n odd) admits a parallel $\frac{1}{2}(n-1)$ -plane Π of nullity $\frac{1}{2}(n-3)$ (case (ii) above). Let the vectors of a normal basis be the null vectors ξ_ρ^i ($\rho \in N$) and the unit vector h^i . The recurrence-formulae are then of the form

$$\xi_{\rho,j}^i = A_{\rho j}^\sigma \xi_\sigma^i + F_{\rho j} e h^i, \quad (6.8)$$

$$h^i_{,j} = -G_j^\sigma \xi_\sigma^i. \quad (6.9)$$

There is no term in h^i on the right-hand side of (6.9) because, if there

were such a term, $H_{\rho} h^{\rho}$ say, we should obtain $H_{\rho} = 0$ on multiplying (6.9) by h_{ρ} and using the fact that h^{ρ} is a unit vector orthogonal to the vectors ξ_{ρ}^i [cf. (6.2) (c)].

If we form a quasi-orthogonal ennuple consisting of ξ_{ρ}^i , h^i and of selected vectors ξ_m^i [$m = \frac{1}{2}(n-1)$] and $\zeta^{\alpha i}$ [$\alpha \in (N, m)$], we again have the general identities (6.2), and these must include (6.8) and (6.9). As before, we separate (6.2) (a) into the two sets (6.4) and (6.5). Comparison of (6.4) and (6.8) shows that $B_{\rho\beta} = 0$ and hence, as before, that

$$B_{\alpha\beta} = 0 \quad [(\alpha, \beta) \in (N, m)], \quad (6.10)$$

and also that

$$A_{\rho j}^m = 0 \quad [\rho \in N, m = \frac{1}{2}(n-1)]. \quad (6.11)$$

Similarly, a comparison of (6.2) (c) and (6.9) shows that

$$F_{\rho j} = 0 \quad [\beta \in (N, m)], \quad (6.12)$$

and that

$$G_j^m = 0 \quad [m = \frac{1}{2}(n-1)]. \quad (6.13)$$

By (6.10), (6.11), (6.12), the first $\frac{1}{2}(n-3)$ equations (6.2) (a), that is, equations (6.4), become (6.8) with $F_{\rho j}$ zero. So we have

$$\boxed{\xi_{\rho, j}^i = A_{\rho j}^{\sigma} \xi_{\sigma}^i} \quad [(\rho, \sigma) \in N]. \quad (6.14)$$

The remaining equation (6.2) (a), namely (6.5), becomes, by (6.10) and (6.12),

$$\boxed{\xi_{m, j}^i = A_{mj}^{\beta} \xi_{\beta}^i} \quad [\beta \in (N, m), m = \frac{1}{2}(n-1)]. \quad (6.15)$$

The first $\frac{1}{2}(n-3)$ of equations (6.2) (b) become

$$\zeta_{\rho, j}^{\alpha i} = C_{j \zeta_{\rho}^{\alpha}}^{\beta} \xi_{\beta}^i - A_{\rho j}^{\sigma} \zeta_{\sigma}^{\alpha i} + G_j^{\sigma} e h^i \quad [\rho \in N, \beta \in (N, m)],$$

and are auxiliary, while the remaining equation (6.2) (b), corresponding to $\alpha = \frac{1}{2}(n-1) = m$, becomes

$$\boxed{\zeta_{\sigma, j}^{m i} = C_{j \zeta_{\sigma}^m}^{\alpha} \xi_{\alpha}^i - A_{mj}^{\alpha} \zeta_{\alpha}^{m i}} \quad [\sigma \in N, m = \frac{1}{2}(n-1)]. \quad (6.16)$$

The first term on the right is $C_{j \zeta_{\sigma}^{\alpha}}^m \xi_{\alpha}^i$ ($\sigma \in N$) and not $C_{j \zeta_{\sigma}^{\beta}}^m \xi_{\beta}^i$ [$\beta \in (N, m)$]

because $C^{\alpha\beta}_j = 0$ on account of the skewness of $C^{\alpha\beta}_j$ in α, β . The second term involves only $\zeta^{\alpha i}$ of the set $\zeta^{\alpha i}$ because of (6.11), and there is no term in h^i because of (6.13).

Finally, (6.2) (c) reduces to (6.9), namely

$$\boxed{h^i_{,j} = -G^i_j \xi^i_\alpha} \quad (\alpha \in N). \quad (6.17)$$

The boxed equations are recurrence-formulae for the case at present under consideration. Equations (6.14) and (6.17) are the recurrence-formulae for the original $\frac{1}{2}(n-1)$ -plane Π of nullity $\frac{1}{2}(n-3)$, of normal basis (ξ^i_ρ, h^i) . Equation (6.14) alone is the recurrence-formula for its null part ω , which is the null parallel $\frac{1}{2}(n-3)$ -plane of basis ξ^i_ρ . Equations (6.14) and (6.15) together show that V_n also admits the null parallel $\frac{1}{2}(n-1)$ -plane, Π_1 say, having $\xi^i_\alpha \equiv (\xi^i_\rho, \xi^i_m)$ as (normal) basis and intersecting Π in its null part ω . Similarly, (6.14) and (6.16) show that $(\xi^i_\rho, \zeta^{\alpha i})$ are a basis, normal because $\zeta^{\alpha i}$ is orthogonal to all the vectors ξ^i_ρ ($\rho \in N$), of another null parallel $\frac{1}{2}(n-1)$ -plane Π_2 , which also intersects Π in ω .

The $\frac{1}{2}(n+1)$ -plane P of basis $(\xi^i_\rho, \xi^i_m, \zeta^{\alpha i})$, which basis is non-normal because $\xi^i_m, \zeta^{\alpha i}$ are not orthogonal, is of nullity $\frac{1}{2}(n-3)$ and is parallel, by (6.14), (6.15), (6.16). It is the conjugate of the original $\frac{1}{2}(n-1)$ -plane Π . Similarly the $\frac{1}{2}(n+1)$ -planes P_1, P_2 , of respective (normal) bases $(\xi^i_\rho, \xi^i_m, h^i)$ and $(\xi^i_\rho, \zeta^{\alpha i}, h^i)$ are of nullity $\frac{1}{2}(n-1)$, and are parallel. They are the conjugates of Π_1, Π_2 . Lastly, the $\frac{1}{2}(n+3)$ -plane p of (non-normal) basis $(\xi^i_\rho, \xi^i_m, \zeta^{\alpha i}, h^i)$ is of nullity $\frac{1}{2}(n-3)$ and is parallel. It is the conjugate of the null part ω of the original plane Π . Clearly p contains all the planes previously considered.

So we have: *If n is odd and V_n admits a parallel $\frac{1}{2}(n-1)$ -plane Π of nullity $\frac{1}{2}(n-3)$, then it also admits the following parallel planes:* (i) the null $\frac{1}{2}(n-3)$ -plane ω , which is the null part of Π ; (ii) two null $\frac{1}{2}(n-1)$ -planes Π_1, Π_2 ; (iii) the $\frac{1}{2}(n+3)$ -plane p , of nullity $\frac{1}{2}(n-3)$, conjugate to ω ; (iv) the $\frac{1}{2}(n+1)$ -plane P of nullity $\frac{1}{2}(n-3)$, conjugate to Π ; (v) the two $\frac{1}{2}(n+1)$ -planes P_1, P_2 , both of nullity $\frac{1}{2}(n-1)$, conjugate to Π_1 and Π_2 . All of these planes intersect in ω and are contained in p .

I have nowhere in this paper considered the conditions of integrability of recurrence-formulae for parallel planes, nor the existence of spaces V_n containing parallel planes of the types discussed. Conditions of integrability for the particular case of a null parallel 2-plane in V_4 were given in my previous paper [equations (5.5), (5.6)], and similar conditions could be

found for the general formulæ (3.5) (a), (c) of this paper. The existence of spaces containing parallel planes of the types considered above follows from theorems due to Walker,* who has obtained canonical forms for the metrics of such spaces.

* A. G. Walker, "Canonical Form for a Riemannian Space with a Parallel Field of Null Planes", *Quart. Journ. Math.* (Oxford, Second Series), 1, 1950, 69-79; and a further paper, at present in the press (May 1950), to appear in the same *Journal*.

(Issued separately June 16, 1950)

VIII.—A Further Note on a Problem in Factor Analysis.*

By D. N. Lawley, M.A., D.Sc., University of Edinburgh

(MS. received November 5, 1949. Read January 9, 1950)

IN a recent paper † by the author several unfortunate omissions and mis-statements occurred which it seems desirable to correct. I am indebted to Professor T. W. Anderson for pointing these out to me.

Firstly, it was stated that the matrix of true loadings K was defined uniquely if it were chosen so as to make $J = K'T^{-1}K$ a diagonal matrix (adopting previous notation). A further condition is however required, namely, that no two elements of J should be equal. This condition would almost certainly be satisfied in practice, but should have been stated explicitly.

It is also necessary to order the columns of K in such a way that the r th column corresponds to the r th largest element of J . The elements of J are then arranged in order of magnitude along the diagonal. A similar consideration applies with regard to the columns of \check{K} , the matrix of estimated loadings. We are thus able to set up a correspondence between the columns of K and those of \check{K} .

A misprint has occurred in § 3 of the paper. Line 3 of page 396 should read as follows:—

$$d_{ij} = 0, \text{ if either } i < m \text{ or } j < m.$$

These restrictions imply the equation

$$KT^{-1}(A - \check{C}) = 0,$$

which, to a sufficient approximation, is equivalent to equation (3).

In § 4 the conditions

$$m_{ir} = 0 \quad (i < r)$$

are equivalent to the assertion that the above-diagonal elements of $K'T^{-1}\check{K}$ are zero, whereas in actual practice \check{K} is chosen such that $\check{K}'T^{-1}\check{K}$ is diagonal. What has in fact been done is to neglect errors of estimation of the loadings in the first $(r-1)$ factors when estimating those in the r th factor. Hence the statement that for $r \neq s$ the covariance between l_{ir} and

* This paper was assisted by a grant from the Carnegie Trust for the Universities of Scotland.

† "Problems in Factor Analysis", *Proc. Roy. Soc. Edin.*, A, LXII, 394-399.

l_{rs} is zero is incorrect. The formulæ (6) represent the *partial* variances and covariances of $l_{1r}, l_{2r}, \dots, l_{nr}$ for given values of the loadings in all previously estimated factors.

While these partial variances and covariances would seem to be of more practical use than the total ones, the values of the latter could probably be obtained by employing previous methods.

(Issued separately June 16, 1950)

IX.—A Measurement of the Velocity of Light.* By **R. A. Houstoun**, M.A., D.Sc., F.Inst.P., Natural Philosophy Department, University of Glasgow. (With Four Text-figures.)

MS. received August 23, 1949. Read November 7, 1949)

SYNOPSIS

The author has already described a new method of measuring the velocity of light, which replaces Fizeau's toothed wheel by a piezo-quartz. This acts as an intermittent diffraction grating, and it interrupts the beam 200 times more rapidly than Fizeau's wheel did.

The present paper describes the application of the method to the measurement of the velocity of light in air. The total length of path was 78 metres. The frequency of interruption was measured by comparison with the Droitwich radio station. The result reduced to vacuum is 299,775 kilometres per second and is in agreement with other recent determinations, but, as a result of the experience gained, it will be possible to increase the accuracy at least ten times.

IN the *Proc. Roy. Soc. Edin.* I have described a new method of measuring the velocity of light and have given results obtained by this method for water (Houstoun, 1941, 1944). The present paper describes an application of the same method to the measurement of the velocity of light in air. Only one previous measurement of the velocity of light in air has been made in this country, that of Young and Forbes † in 1881, the eleven determinations considered by Birge (Birge, 1941) in his review of the subject being made in America, France and Germany. So this paper breaks a long silence.

METHOD

As has been explained in the previous papers, my method consists in replacing Fizeau's toothed wheel by a rectangular piezo-quartz. This has the property of acting as an intermittent diffraction grating when it is placed in an alternating electric field. Stationary elastic waves are set up in the quartz, the refractive index being increased at condensations and

* The research was assisted and this paper was assisted in publication by grants from the Carnegie Trust for Universities of Scotland.

† George Forbes, it is interesting to note, was an F.R.S.E. and Professor of Natural Philosophy at the Andersonian College in Glasgow, and his determination was made across the Firth of Clyde from Wemyss Bay to Innellan. His father was Professor of Natural Philosophy in the University of Edinburgh, and Principal of the University of St Andrews.

diminished at rarefactions; thus the surface of a light-wave travelling in the quartz at right angles to the elastic wave becomes corrugated, and these corrugations cause diffracted beams to appear on each side of the main beam twice in each electrical period.

Let SQ (fig. 1) represent a beam entering the quartz and QM a first diffracted beam. If this is reflected back on its path by a mirror M and arrives at Q when the grating is in action, it is diffracted back again to S; otherwise it travels on to S'. The advantage of this method of interrupting

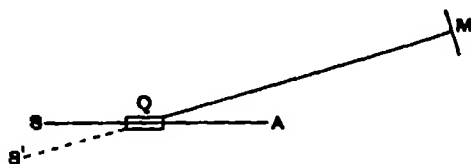


FIG. 1.

the beam over the toothed wheel is that, as I have used it, the beam is interrupted 200 times as rapidly. There is, however, one important difference. The speed of the wheel is varied to make the intensity of the returning beam a minimum. But the quartz behaves as a grating only when one of its free periods is in resonance with the alternating field; thus the minimum can be obtained only by sliding M along QM.

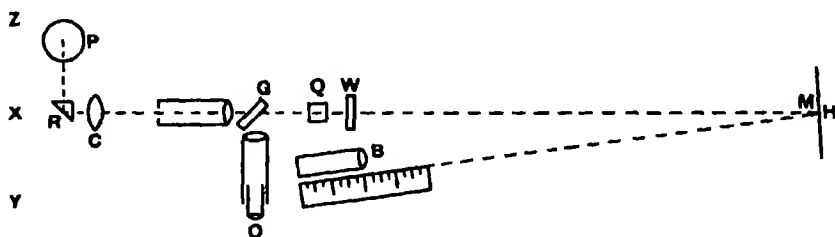


FIG. 2.

Fig. 2 shows the experimental arrangement used for the present measurement. P is a 100-candle-power Pointolite lamp, which was enclosed in a box; the light was reflected by a 45° glass prism and focused by the condenser C on the slit of a spectrometer. After being rendered parallel the beam passed through the glass plate G and the piezo-quartz Q, which lay between the condenser plates of an oscillator, not shown in the figure. It then travelled a distance of about 19 metres to the plane mirror M, was reflected to the lens-mirror combination B, returned on its path, and was reflected by the glass plate G through the telescope to the eye of the observer at O.

The deviation of the first-order spectrum is about $35'$. The adjustment

of the mirror M and lens-mirror B is done best with the direct image. The end of the quartz was bevelled, therefore, so as to give the direct image a deviation of $35'$, and a glass wedge W inserted to remove this deviation. The adjustment of the line was done with this wedge in position. When the oscillator was started and the wedge removed, the first-order spectrum automatically used the same line.

The oscillator is represented diagrammatically in fig. 3. It is a push-pull one employing two Osram D.E.T. 12 valves, and the quartz was placed between the plates of the condenser C. These were horizontal; their surfaces measured 40×13 mm.³ and they were about 1 mm. thick.

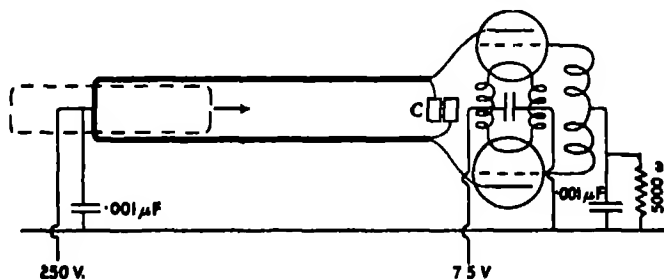


FIG. 3.

The quartz measured $3.51 \times 11.03 \times 12.975$ mm., and the upper condenser plate rested loose on the top of it. The 11.03 edge was in the direction of the optic axis. The parallel rods are 26 cm. long and 4.5 cm. apart. Tuning was done by moving a copper rectangle backwards and forwards by a screw motion. The position of this rectangle is indicated by the broken line; it was about 8 mm. above the plane of the parallel rods. Filament current was supplied by four large storage cells, and the anode current was taken from two 120-volt high tension batteries. The observer had the slow motion for tuning the oscillator close to his right hand, and its frequency range was from about 104 to about 110 megacycles per second. The distance GQ was about 40 cm.; this was necessary to allow stray light from the quartz to diffuse away before reaching the object-glass of the telescope.

The apparatus was set up on two tables in a research room on the second floor of the department, and the beam left this room by a window and travelled the full length of a students' laboratory to the mirror M; measurements could be made only in the vacation when the laboratory was not in use. The mirror was a heliograph one, of diameter 13 cm.

The length of the path was measured by a steel tape supplied by James Chesterman & Co., Leeds; this hung in a single catenary under a stretching force of 20 lb. It was correct at 68° F., and its coefficient of

expansion was 6.2×10^{-6} per degree Fahrenheit. The one end was hooked on a pillar at H; the other hung either over a cylinder at X or over a cylinder at Y, carrying weights at its end, and it could be lifted easily from the line XH to the line YH, or out of the way to the line ZH. The tape had only two graduations, at 0 and at 20 metres, and the positions of the quartz, mirror M and lens-mirror B were obtained by fixing 100 and 20 cm. graduated metal scales on to it and measuring from these points.

The lens-mirror combination B consisted of an achromatic lens of 25 cm. focal length and 4 cm. diameter, and a thin glass mirror mounted at its focus. They were fixed in a tube the direction of which could be adjusted very accurately, and which slid along a metal optical bench of length 110 cm.

The measurement of the velocity falls into two parts: the measurement of the electrical wave-length—*i.e.* the distance light travels in air during one electrical period—and the measurement of the electrical frequency. After the mirror M and the lens-mirror B were aligned, the observer tuned the oscillator into resonance with an harmonic of the quartz and removed the wedge W. This displaced the first-order image into the position previously occupied by the direct image. He then adjusted the lens-mirror, so as to make the intensity of the image a minimum. The lens-mirror was close to his right hand when his eye was at the eyepiece. Twenty settings were made in succession and the mean taken. The electrical wave-length was derived from this mean in the following manner.

The tape was switched into the line XH and the positions of the right end of the quartz and the glass-air face of the mirror noted. The lens-mirror was then moved to the end of its scale, the tape switched to the line YH, and the positions of the end of the tube of the lens-mirror and the glass-air surface of M again noted. The tape passed about 1 cm. above the top of M, about 5 cm. above the quartz and about 1.8 cm. above the end of B. The readings were taken by laying a straight edge on the face of the mirror, and by placing a celluloid square on the top of B or on the surface on which the condenser plates rested. They could be made almost to one-tenth of a millimetre. They gave the uncorrected length.

Corrections have to be made for (a) half the optical length of the quartz, (b) the optical thickness of the mirror M, (c) the object-glass of the lens-mirror, and (d) the distance from the end of the tube of the lens-mirror to the reflecting surface. Their values are 0.87, 0.91, 0.27 and 1.89 cm. respectively, giving a total of 3.94 cm. The group index of refraction was used, but the difference is hardly appreciable.

It will make matters clearer if we give the numerical values for one

set of readings, the twelfth set, selected because they were average values. The twenty observations in this set gave a mean scale reading of 19.45 cm. The uncorrected length was 3916.9 cm. This makes the corrected length

$$3916.9 + 3.94 - 19.45 = 3901.4 \text{ cm.}$$

As the light traverses this going and coming, we have finally for the length of the path

$$7802.8 \text{ cm.}$$

For a minimum this must equal $(n + \frac{1}{2})\frac{\lambda}{2}$, where n is an integer. For the frequency used n was 55. Thus, in this case,

$$\lambda = \frac{7802.8 \times 4}{111} = 281.18 \text{ cm.}$$

The frequency was measured by a wavemeter consisting of an oscillator working through the range 8 to 12 megacycles per second, and a one-megacycle per second quartz crystal oscillator, which controlled multivibrators at frequencies of 100 and 10 kilocycles per second. For the measurements recorded in the present paper my piezo-quartz was operated on its 135th harmonic. Its frequency was then slightly greater than 13×8200 kc., but during a set of readings the quartz heated up and the frequency generally fell below this value. No attempt was made to keep the temperature constant; the change was small and occurred at a uniform rate, so it could be eliminated mathematically. It was found advantageous to keep the 10 kc. multivibrator permanently switched off and measure the frequency of the piezo-quartz from the 8200 kc. note. There was no difficulty in detecting the one note from the other; the 100 kc. multivibrator gave a slow and leisurely resonance, the piezo-quartz one that was louder and much sharper.

During the twelfth set of observations the piezo-quartz resonance moved from the multivibrator resonance to $\frac{1}{10}$ division of the condenser scale below it. A division on this scale is equal to 1.96 kc. The mean frequency of the piezo-quartz was thus

$$(8200 - \frac{1}{10} \times 1.96) \times 13 = 106,596 \text{ kc.}$$

These observations therefore gave for the velocity of light in air

$$281.18 \times 1.06596 \times 10^8 = 2.9973 \times 10^{10} \text{ cm. per sec.}$$

The wavemeter was standardised by means of the Droitwich carrier wave; the frequency of the latter is 200 kc. per sec., and it is maintained very closely to this value. It was received by an aerial, amplified, and put

on the X plates of a Double Beam Cossor oscillograph. The 100 kc multivibrator was put on the Y plates. The pattern was usually almost stationary, and never rotated faster than once in three seconds, so it was never necessary to adjust the frequency of the one megacycle crystal.

RESULTS

In the final determination 400 settings were made. These were made in 20 sets, 20 in each set, during the month of August 1949. They were consecutive settings; once a setting was made, the reading was entered on the record, no matter how "bad" it was, and once a set was started, it was completed. If the work did not appear to be going well, longer rests were taken between the settings. I made all the settings myself. The frequency was taken four times during each set, temperature and barometer were read after each set, all the distances were measured after each set, and everything was thrown completely out of adjustment and readjusted before the next set was started. The results for each set are given in 10^6 cm. per sec. in the following table; they are for air, and they are corrected for the expansion of the measuring tape.

1	29961	11	29966
2	29968	12	29973
3	29979	13	29974
4	29961	14	29972
5	29979	15	29975
6	29976	16	29965
7	29973	17	29977
8	29959	18	29962
9	29972	19	29972
10	29963	20	29969

They are graphed in fig. 4. The points in this graph represent the mean results obtained from the first 20, 40, 60, 80 . . . observations, and show the approach of the mean towards its final value. In the review already cited Birge makes an estimate of the value of the velocity on the basis of the most accurate determinations hitherto made, and in a more recent critical examination of the field, Dorsey (Dorsey, 1944) makes a similar estimate. Birge's and Dorsey's means, reduced to air, are shown respectively by the upper and lower lines in the graph, and my final point lies between them, though, the probable error being what it is, this must be regarded as an accident.

My final value for air is $299,698 \pm 9$ km. per sec. Using Birge's correction, this becomes $299,775 \pm 9$ for a vacuum. The probable error was calculated from the 20 means given in the preceding table. The following table gives other values for a vacuum:—

Birge, estimated	299776 \pm 4	} As revised by Birge
Dorsey, estimated	299773 \pm 1	
Karolus and Mittelstaedt, 1929	299786 \pm 10	
Michelson, Pease and Pearson, 1935	299774 \pm 4	
Anderson, 1937	299771 \pm 10	
Hüttl, 1940	299771 \pm 10	
Anderson, 1941	299776 \pm 6	

Michelson, Pease and Pearson's determination is the mean of 2885 observations, Anderson's second determination of 2895 observations.

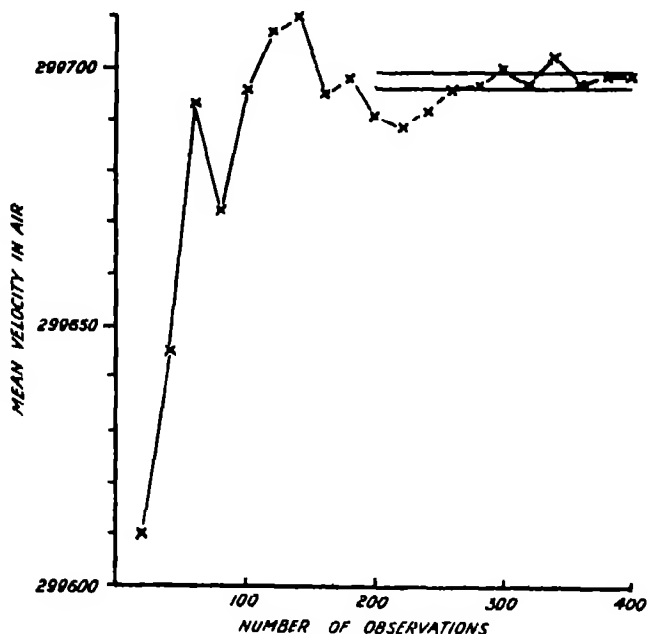


FIG. 4.

An interesting point has cropped up in connection with the probable error. The observer sits with his chin on a rest and his eye at the telescope and sweeps the lens-mirror back and forward along the bench, until he is going from just perceptible increment on the one side to just perceptible increment on the other. He then crosses this region in little steps, usually six, and, when he thinks he is doing this satisfactorily, stops at the third step. This should bisect the difference. But if he has the habit of stopping too soon or too late, the observations fall into two groups according as the final steps were inward or outward, and each of these groups has its own probability curve. The errors of the one group are to some extent compensated for by the errors of the other. If this is disregarded, the calculated probable error will be too high. In my measurement of the value of the velocity of light in water (Houstoun, 1944) I comment on the fact

that the agreement with theory is much better than would be expected from the probable error. In that determination the settings were made "in" and "out" alternately, and I have now no doubt that the value given for the probable error is too high. In the present investigation there were 200 "ins" and 200 "outs". They were made in no definite order but just as they came, only the last six or seven being directed, in order to make the totals balance. Although some of the 20 sets were well balanced, others were predominantly "in" or "out". Unfortunately the bearing of this point on the probable error did not occur to me until after the paper recording the directions of the single settings was thrown away. So I calculated the probable error, as if the readings were independent, and the value given, 9 km. per sec., is definitely too high.

I have not seen the above point referred to before; it may affect the estimates of the accuracy of some of the other determinations of the velocity of light.

I have been asked whether the accuracy of the later measurements is greater than that of the earlier, and whether a better mean could be obtained by neglecting some of the earlier. The later measurements are slightly nearer the mean, but that by itself is not a reason for preferring them, and I think it is safer, as I have done, to give all observations equal value. Later measurements are sometimes more consistent owing to habit-forming; in the present case this was guarded against by displacing the mirror M an inch or two between the sets of readings.

The resonance bands of the piezo-quartz are about one kilocycle wide. The low frequency edge is sharp and very bright, and the intensity diminishes gradually from this edge to the other side. The sharpness of the minimum depends on the intensity, there being one particular frequency which gives the best results.

During the measurements the temperature of the laboratory varied from 64 to 80° F. and the barometer from 75.4 to 76.2 cm. It was not necessary to apply a correction for pressure. The average correction for the temperature of the measuring tape was about 2 km. per sec.; the maximum correction for the same was 19 km. per sec.

DISCUSSION OF THE RESULTS

The chief source of error lies in determining the position of the minimum. Various optical devices were tried to diminish this, but they were all found inferior to the direct method. Some observers are much better at determining the minimum than others; I am above the average, but I have had one observer who could fix it three times as accurately as I can.

With a powerful arc lamp it would be possible to increase the range ten times. So with the same apparatus and procedure the accuracy could be increased thirty times.

The accuracy might also be increased by increasing the frequency of the oscillator, as it is the ratio of electrical wave-length to length of path that matters. I worked on an electrical wave-length of 281 cm. If this were diminished to 28.1 cm., with the same range the accuracy would be increased ten times. It has yet to be shown, however, that a quartz would give bright diffraction bands on this wave-length.

For interrupting the beam, Mittelstaedt, Anderson and Hüttl used a Kerr cell in an alternating field. Mittelstaedt received the beam through another Kerr cell, which was placed in a field of the same frequency, and obtained his minimum by altering the frequency. At his minimum the intensity went down to zero. Anderson used two paths of unequal length. After passing through the Kerr cell the light divided at a mirror, and the two beams reunited at the same mirror and fell on a photoelectric multiplier. The current from the multiplier passed through a circuit tuned to the frequency of the interruption. If the groups arriving by the two paths were out of phase, the response was a minimum. The principle is thus that of the neutralising condenser. It is arranged that the groups are out of phase by drawing back a mirror; the difference of path is then an odd multiple of half the electric wave-length. Hüttl received his beam on a vacuum photo-cell, the anode potential of which varied with the same frequency as the Kerr cell. Anderson worked at 19.2 megacycles per second and Hüttl at 4-12 megacycles per second, and it is doubtful whether their techniques would work at my frequency. There might be trouble with the transit times of the electrons in the photo-cells. This is a point that should be investigated.

Rotating mirrors, toothed wheels and rotating prisms are now definitely outclassed; they are too sensitive to the friction of their bearings, and they require great distances if the result is to be accurate. No one working on these lines will do better than Michelson, Pease and Pearson. The Kerr cell must be water-cooled, and requires a voltage of 6000 volts or so to bias it and an alternating e.m.f. of about 2000 volts to operate it. It seems to have reached its limiting frequency, it is somewhat messy to work with, and its possibilities have been fully explored. The requirements of the piezo-quartz are more modest and it gives a feeling of æsthetic satisfaction. There are no systematic errors or concealed traps associated with it. But the minimum is unsatisfactory, having about one-third the intensity of the maximum, and to say that its determination gives no feeling of æsthetic satisfaction is an understatement. In any case, either of these methods

is capable of giving a value for the velocity at least ten times as accurate as anything that has hitherto been obtained.

A length can be measured by the time that light takes to traverse it, and it is interesting to note that this method is at present being developed in Sweden (Bergstrand, 1943) for use by the Ordnance Survey in measuring the sides of triangles. The apparatus is based on Hüttl's method.

ACKNOWLEDGMENTS

In making the frequency measurements recorded in this paper I have received much help from Dr Walter McFarlane; I have also received assistance in tuning up the apparatus from two of my students, Mr G. H. Robertson and Mr J. I. B. Macfadyen.

REFERENCES TO LITERATURE

- BERGSTRAND, ERIK, 1943. "Measurement of Distances by High Frequency Light Signalling", *Ark. Mat. Astr. Fys.*, xxixA, N: o 30.
- BIRGE, RAYMOND T., 1941. "The General Physical Constants", *Reports on Progress in Physics*, VIII, 90.
- DORSEY, N. ERNEST, 1944. "The Velocity of Light", *Trans. Amer. Phil. Soc.*, xxiv, part 1.
- HOUSTOUN, R. A., 1941. "A New Way of Measuring the Velocity of Light", *Proc. Roy. Soc. Edin.*, A, LXI, 102.
- , 1944. "A Measurement of the Velocity of Light in Water", *ibid.*, A, LXII, 58.

(Issued separately June 16, 1950)

X.—The Reciprocity Theory of Electrodynamics.* By **H. S. Green** and **K. C. Cheng**, University of Edinburgh. *Communicated by Professor MAX BORN, F.R.S.*

(MS. received July 1, 1949. Read December 5, 1949)

SYNOPSIS

This paper represents the application of the Principle of Reciprocity, formulated in a previous communication, to the outstanding problems of classical and quantum electrodynamics.

The first step consists in the formulation of a reciprocally invariant Lagrangian function for a system of electrons in interaction with the electromagnetic field. A study is made of the unaccelerated motion of an electron, and this is subsequently extended to embrace the problem of an electron in arbitrary motion. It is found that the usual difficulties of classical electrodynamics do not appear. The methods of the earlier paper are applied to the derivation of the Hamiltonian energy of electron and field, and this enables a quantized formulation of the theory to be given, which also does not lead to the usual divergence difficulties.

1. INTRODUCTION

IT is well known that the present state of the theory of the electron, in both classical and quantum electrodynamics, cannot be regarded as satisfactory owing to certain difficulties of a fundamental nature. In the classical theory these difficulties are associated with the divergence of the self-energy of the electron when it is regarded as a point, and the failure of various alternative models with a finite structure to meet relativistic requirements. Early hopes that the application of the quantum theory would remedy these defects were discouraged when it became apparent that the very process of quantization introduced divergent terms additional to those already present in the classical theory. Even in the "hole theory" of Dirac the calculations of Weisskopf (1934) revealed a divergence of both the electrostatic and electromagnetic self-energy of the electron, which further recent developments by Schwinger (1948, 1949) among others have not removed.

Naturally very many devices for the elimination of the divergence difficulties have been suggested. The most direct method available consists in the subtraction of the divergent integrals in the way proposed by Dirac (1938, 1942) and Pryce (1938), or by Heitler and Peng (1942). The objection which now has to be met by all of these "subtraction" theories is that where the elimination of the divergences is not incomplete,

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

no account can be given of phenomena such as the recently discovered "Lamb effect", which by common consent must be explained with the help of the disappearing terms.

Various other methods may be distinguished and are summarized in an excellent review by Pais (1948), which includes a very full bibliography. It is clear that the necessary departure may be made in two main directions: either one may introduce non-linearity in the radiation field, as attempted by Born and Infeld (1934), or, without modifying the linear nature of the field equations, various procedures, *e.g.* those suggested by Bopp (1940, 1943), Podolsky (1942) and McManus (1948), all equivalent to the introduction of multiple derivatives of the field variables into the Lagrangian, may be employed.

The method here adopted has its genesis in the theory of reciprocity proposed by Born (1939), a development of which was made recently, in association with the present authors, in a paper (Born, Green *et al.*, 1949) here called Q.T.R. It is first necessary to specify the form of the wave operator $F_{kl}(x, x')$ introduced in Q.T.R. appropriate to the electromagnetic field; the field equations are then easily obtained by the general method which is equivalent to the application of the principle of least action. The electrons are represented as point singularities in the electromagnetic field, and a reciprocally invariant Lagrangian is derived from which the equation of motion of the electron can be obtained by variation with respect to the electronic co-ordinates.

The resulting field equations are readily solved for the motion of an unaccelerated singularity, in a relativistically invariant way which introduces the reciprocally invariant four-dimensional angular momentum tensor. For accelerated motion it is also possible to obtain an exact formal solution which can be shown to coincide with Wiechert's solution of Maxwell's equations at large distances r from the electron, and for short distances can be expanded in powers of r and the fundamental length a . The result is finite everywhere, and to a first approximation reduces to the difference between the advanced and retarded solutions of Maxwell's equations for the field of a point singularity as postulated by Dirac. The equation of motion of the electron can then be derived, also as a power series in a , and in first approximation it is identical with that obtained by Dirac. There is, however, an important deviation from Dirac's theory in the form of the interaction between two electrons at short distances; this indicates that it should be possible to obtain a classical theory of pair creation and annihilation without the divergent terms of the usual theory.

As a preliminary to the quantization of the theory, a Hamiltonian

formulation is essential; this may be obtained by an extension of the theory developed for pure fields in Q.T.R. It is easy to write down an exact expression for the energy; unfortunately this is not of the type to which ordinary methods of quantization are readily applied, as it contains high derivatives of the field variables with respect to time. It would not be correct to treat these derivatives as independent variables, and they must therefore be eliminated; this is accomplished with the help of the field equations.

In order to apply the usual perturbation methods of quantum mechanics, it is necessary to separate from the Hamiltonian a term which may legitimately be regarded as small. Any term involving the self-field of the electron does not satisfy this requirement, and it therefore has to be separated from the resultant field—a procedure analogous to the separation of the longitudinal field in orthodox electrodynamics. The resulting theory is equivalent to the usual one apart from the presence of a factor $\exp \{(\mathbf{k} \cdot \mathbf{p}/E)^2 - k^2\}/4b^2$ in the interaction energy, where \mathbf{p} and E are the momentum and energy of the electron before the emission, or after the absorption, of a photon with momentum \mathbf{k} . This rule, secured by a proper application of the correspondence principle, ensures that the contribution of all the "round-about" transitions to the self-energies and cross-sections are small and finite, while leaving the real processes unaffected, except to a small and rather interesting degree which, it may be hoped, improvements in experimental technique may confirm.

2. LAGRANGIAN FORMULATION

In the original paper Q.T.R. (Born, Green *et al.*, 1949) a general principle was established according to which the Lagrangian density of any pure field must have the form

$$\frac{1}{2}F(p, p')\rho(x, x'), \quad (2.1)$$

where $F(p, p')$ is a reciprocally invariant operator satisfying the condition

$$(2\pi\hbar)^{-3} \int F(p, p) \exp(ip_k x^k/\hbar) dp = F(x, x) \quad (2.2)$$

provided the variables p_k and x_k are regarded as imaginary. This principle has been shown to lead to the prediction of a set of rest-masses for the elementary particles in sufficient agreement with those experimentally observed in cosmic radiation. The purpose of the present paper is to examine the further application of the "principle of reciprocity" to classical and quantum electrodynamics.

For the pure electromagnetic field, $\rho(x, x')$ is of the form

$$\rho_{kl}(x, x') = A_k(x)A_l(x'), \quad (2.3)$$

where k and l are tensor affixes subject to the conventions of general relativity theory but referring to the Galilean metric

$$g_{11} = g_{22} = g_{33} = -1, \quad g_{44} = 1, \quad X_k \equiv (x, ct). \quad (2.4)$$

The components of the four-vector $A_k \equiv (A, \phi)$ are then the potentials of the electromagnetic field, and for such a vector field containing only particles with vanishing rest-mass one has

$$F_{kl} = f(P)(-Pg_{kl} + cp_k p_l), \quad P = p^k p_k = -\hbar^2 \square, \quad (2.5)$$

where c is a disposable constant, and $f(P) = 0$ may have no real roots. The further condition that F_{kl} should be self-reciprocal determines the solution uniquely, with

$$c = 4, \quad \hbar^2 f(P) = \exp(-P/2b^2) \quad (2.6)$$

obtained by substituting $k=2$ and $n=3$ in the equations of § 5 of Q.T.R. Then the field equations for the pure electromagnetic field, corresponding to Maxwell's equations in orthodox electrodynamics, are

$$-\hbar^{-2}(P\delta_l^k - 4p^k p_l)A_k(x) \exp(-P/2b^2) = 0. \quad (2.7)$$

The four-vector A_k is specified completely only when the auxiliary condition

$$\frac{\partial}{\partial x_k} \cdot A_k = 0 \quad (2.8)$$

is supposed to hold identically; this condition is introduced not of necessity but in order to simplify the subsequent calculations. Then (2.6) reduces to $\exp(-P/2b^2) \cdot P A_k(x) = 0$, and this can be further simplified by the observation that there is no non-vanishing solution of the equation $\exp(-P/2b^2)\chi = 0$. For let

$$\chi(x) = (2\pi\hbar)^{-3} \int \omega(p) \exp(ip_k x^k / \hbar) dp^{(4)}, \quad dp^4 = dp dp_4;$$

then, by definition,

$$\exp(-P/2b^2)\chi(x) = (2\pi\hbar)^{-3} \int \omega(p) \exp(-p_k p^k / 2b^2 + ip_k x^k / \hbar) dp^{(4)},$$

which vanishes only if $\omega(p) = 0$ identically. From this it follows that

$$P A_k(x) = -\hbar^2 \square A_k(x) = 0, \quad (2.9)$$

which is equivalent to Maxwell's equations.

The energy and momentum four-vector P_m of the field, derived straightforwardly in accordance with the principles explained in Q.T.R., may be obtained as follows:—

$$\begin{aligned} F_{kl}(\rho, \rho') &= \hbar^{-2} e^{-(P+P')/4\hbar^2} \{ -\dot{\rho}_m \rho^m g_{kl} + 2(\rho_k \dot{\rho}_l + \dot{\rho}_k \rho_l) \}, \\ G_{kl}^m(\rho, \rho') &= \hbar^{-2} e^{-(P+P')/4\hbar^2} \{ -\dot{\rho}^m g_{kl} + 2(\delta_k^m \dot{\rho}_l + \dot{\rho}_k^m \rho_l) \\ &\quad + (\dot{\rho}_m + \rho_m) \left(\frac{e^{(P-P')/4\hbar^2} - 1}{P' - P} \right) (-\dot{\rho}^m g_{kl} + 4\dot{\rho}_k \rho_l) \}; \end{aligned} \quad (2.10)$$

$$\begin{aligned} P_m &= \frac{1}{2} \int \langle (-F_{kl} \delta_m^4 + G_{kl}^m \rho_m + G_{kl}^m \dot{\rho}_m) A^k A^l \rangle dx^{(3)} \\ &= \frac{1}{2} \int \left(-\frac{\partial B^k}{\partial x_l} \frac{\partial B_k}{\partial x^l} \delta_m^4 + 2 \frac{\partial B^k}{\partial x_4} \frac{\partial B_k}{\partial x^m} \right) dx^{(3)}, \end{aligned} \quad (2.11)$$

where the pseudo-potentials B_k are defined by

$$B_k(x) = e^{-P/4\hbar^2} A_k(x). \quad (2.12)$$

It is clear from these expressions that the theory of the pure electromagnetic field is indistinguishable from the orthodox theory provided A_k is replaced by B_k .

When there is a system of electrons in interaction with the electromagnetic field, it is assumed that the field equations (2.7) remain valid except at certain singular points which may be interpreted as the positions of the electrons. Then *

$$\begin{aligned} e^{-P/4\hbar^2} \square A^k(x) &= e_0 \sum_i \delta(\mathbf{x} - \mathbf{x}_{(i)}(t)) \dot{x}_{(i)}^k / c \\ &= e_0 \sum_i \int_{-\infty}^{\infty} \delta(\mathbf{x} - \mathbf{x}_{(i)}(s)) \delta(t - t_{(i)}(s)) \frac{dx_{(i)}^k}{ds} ds, \end{aligned} \quad (2.13)$$

where the first line can be obtained from the obviously Lorentz-invariant second line by expressing the proper time s of the electrons as functions of $t_{(i)}$, and performing the integrations with respect to $t_{(i)}$. These field equations can be derived from the Lagrangian

$$L = \int \{ \frac{1}{2} F_{kl} A^k A^l - A^k R_k \} dx^{(3)} + \sum_i L_{(i)}, \quad (2.14)$$

where R^k stands for the right-hand side of (2.13), and $L_{(i)}$ is a function of the electronic co-ordinates and their time derivatives only.

* A dot placed above a symbol, thus: \dot{x} , here and elsewhere, denotes differentiation with respect to time. $e_0 = \sqrt{4\pi}e$ is the electronic charge in Heaviside units, which are used throughout this paper.

Let $\sum_i A_{(i)}^k$ be the particular solution of the equation (2.13), from which the general solution A_i^k may be obtained simply by the addition of the complementary solution A_r^k , which represents a pure electromagnetic field, thus

$$A^k = \sum_i A_{(i)}^k + A_r^k. \quad (2.15)$$

Then the Lagrangian (2.14) can be expressed in the form

$$L = \frac{1}{2} \int \langle F_{kl} \{ A^k A^{l'} - \sum_i (A_{(i)}^k A^{l'} + A^k A_{(i)}^{l'}) \} \rangle dx^{(3)} + \sum_i L_{(i)},$$

as can be seen by the substitution of $\sum_i e^{-p/2\delta} \square A_{(i)}^k$ for R^k . It will be shown in the following section that $L_{(i)}$ can be expressed in the form

$$L_{(i)} = -\frac{1}{2} \int \langle F_{kl} A_{(i)}^k A_{(i)}^{l'} \rangle dx^{(3)}. \quad (2.16)$$

This expression is obviously a function of the electronic co-ordinates and their derivatives only, and will be shown to reduce to

$$L_{(i)} = -m_0 c^2 (1 - \frac{1}{2} \frac{v^2}{c^2})^{\frac{1}{2}}, \quad m_0 = \frac{e_0^2}{2(2\pi)^{\frac{1}{2}} a c^2}, \quad (2.17)$$

which is the generally accepted Lagrangian for a free particle with mass m_0 . The entire Lagrangian may then be written in the form

$$L = \frac{1}{2} \int \langle F_{kl} \{ A^k A^{l'} - \sum_i (A_{(i)}^k A^{l'} + A_{(i)}^k A^{l'} + A_{(i)}^k A_{(i)}^{l'}) \} \rangle dx^{(3)}, \quad (2.18)$$

which is required by the Principle of Reciprocity.

It may be remarked that although the preceding theory is used throughout subsequent sections of the present paper, there exists another formulation in which $F(p, p')$ is regarded as self-reciprocal provided the condition

$$(2\pi\hbar)^{-4} \iint F(p, p') \exp i(p_k x^k - p'_k x'^k) / \hbar \cdot d\rho^{(4)} d\rho'^{(4)} = F(x, x') \quad (2.19)$$

is satisfied, instead of (2.2). This leads to the formula

$$F(p, p') = e^{-(p+p')/2\delta} \{ p_k p'^k \delta_b^l - 2(p_k p^{l'} + p'_k p^l) \} \quad (2.20)$$

instead of (2.10). It can be seen, therefore, that the consequences of the two formulations of the theory are identical apart from the value of δ , which, assuming (2.20), would have a value $\sqrt{2}$ times the value implied by (2.10).

3. UNACCELERATED AND UNIFORMLY ACCELERATED MOTION OF ELECTRONS

The solution of the equation (2.13) is readily obtained for an electronic singularity \mathbf{x} , whose path is a straight line, so that

$$\exp(-P/2b^2) \square A_k(x) = e_0 V_k \delta(\mathbf{r})/c, \\ \mathbf{r} = \mathbf{x} - \mathbf{x}_0 = \mathbf{x} - \mathbf{x}_0 - \mathbf{v}(t - t_0), \quad (3.1)$$

where \mathbf{x}_0 is the position at some initial time t_0 . It will be convenient to obtain the solution of the somewhat more general equation

$$F(P)\phi(x) = e_0 \delta(\mathbf{r}). \quad (3.2)$$

By definition one has

$$F(P)\phi(x) = (2\pi\hbar)^{-3} \int F(p_k p^k) \psi(p) \exp(ip_k x^k/\hbar) dp^{(4)}, \\ \psi(P) = (2\pi\hbar)^{-3} \int \phi(x) \exp(-ip_k x^k/\hbar) dx^{(4)}. \quad (3.3)$$

Also

$$(2\pi\hbar)^{-3} \int \delta(\mathbf{r}) \exp(-ip_k x^k/\hbar) dx^{(4)} \\ = (2\pi\hbar)^{-1} \exp[ip \cdot (\mathbf{x}_0 - \mathbf{v}t_0)/\hbar] \cdot \delta(p_4 - \mathbf{v} \cdot \mathbf{p}/c). \quad (3.4)$$

Hence

$$F(p_k p^k) \psi(p) = \frac{e_0}{2\pi\hbar} \delta(p_4 - \mathbf{v} \cdot \mathbf{p}/c) \exp[ip \cdot (\mathbf{x}_0 - \mathbf{v}t_0)/\hbar]; \\ \phi(x) = (2\pi\hbar)^{-3} \int \psi(p) \exp(ip_k x^k/\hbar) dp^{(4)} \\ = e_0 (2\pi\hbar)^{-3} \int \exp(-ip \cdot \mathbf{r}/\hbar) / F\{(\mathbf{v} \cdot \mathbf{p}/c)^2 - \mathbf{p}^2\} d\mathbf{p}^{(3)}. \quad (3.5)$$

The x_1 -axis may be chosen in such a way that it coincides with the vector \mathbf{v} ; then, writing

$$\gamma = (1 - v^2/c^2)^{-1/2}, \quad \mathbf{k} = (p_1 \gamma^{-1}, p_2, p_3), \quad k = |\mathbf{k}|, \\ \mathbf{u} = (\gamma v_1, v_2, v_3), \quad u = |\mathbf{u}|, \quad (3.6)$$

it follows that

$$\phi(x) = e_0 \gamma (2\pi\hbar)^{-3} \int \exp(-i\mathbf{k} \cdot \mathbf{u}/\hbar) / F(-k^2) d\mathbf{k} \\ = 2e_0 \gamma u^{-1} (2\pi\hbar)^{-3} \int_0^\infty \sin(ku/\hbar) k / F(-k^2) dk. \quad (3.7)$$

With $F(P) = -\hbar^{-2}P \exp(-P/2b^2)$, as required by (3.1), one obtains

$$\begin{aligned} A_k(x) &= e_0 \gamma v_k (ac)^{-1} Y(u), \\ Y(u) &= 2e_0 (2\pi)^{-2} u^{-1} \int_0^\infty \sin(ku/\hbar) k^{-1} \exp(-k^2/2b^2) dk \\ &= a(2\pi)^{-2/2} u^{-1} \int_0^{u/a} \exp(-\frac{1}{2}y^2) dy, \quad (a = \hbar/b). \end{aligned} \quad (3.8)$$

The pseudo-potentials B_k defined by (2.12) satisfy

$$\exp(-P/4b^2) \square B_k = e_0 v_k \delta(r), \quad (3.9)$$

and are accordingly given by

$$\begin{aligned} B_k(u) &= e_0 \gamma v_k (ac)^{-1} Z(u), \\ Z(u) &= 2a(2\pi)^{-2} u^{-1} \int_0^\infty \sin(ku/\hbar) \exp(-k^2/2b^2) \frac{dk}{k} \\ &= a\pi^{-2/2} (2u)^{-1} \int_0^{u/a} \exp(-y^2) dy. \end{aligned} \quad (3.10)$$

The scalar magnitude u is a relativistic invariant, as may be shown by expressing it in terms of the four-dimensional angular momentum tensor, defined by

$$m^{kl} = (x^k - x_0^k) p_0^l - (x^l - x_0^l) p_0^k, \quad p_0^k = m \gamma v^k. \quad (3.11)$$

The scalar of this tensor is

$$\begin{aligned} \frac{1}{2} m^{kl} m_{kl} &= m^2 c^2 \gamma^2 \{ \mathbf{r}^2 - (\mathbf{r} \wedge \mathbf{v})^2 / c^2 \} \\ &= m^2 c^2 u^2. \end{aligned} \quad (3.12)$$

If $t = t_0 + |\mathbf{x} - \mathbf{x}_0|/c$, u reduces to the "retarded distance" of the point (\mathbf{x}, t) from the path of the singularity. For large u , both A_k and B_k given by (3.8) and (3.10) reduce to $e_0 \gamma v_k (4\pi u)^{-1}$, and accordingly coincide with the generally accepted potentials of the electron.

The solution (3.8) may be substituted in (2.16) to obtain a first approximation (with neglect of acceleration) to the exact Lagrangian of an electron in interaction with the pure electromagnetic field. With the help of (3.12) one has

$$\frac{\partial}{\partial x^k} A_e^k = e_0 \gamma v_k (ac)^{-1} \frac{\partial Y}{\partial u} \frac{\partial u}{\partial x^k} = e_0 \gamma v_k (ac)^{-1} \frac{\partial Y}{\partial u} m_{kl} \gamma v^l / mc^2 = 0, \quad (3.13)$$

so that the condition (2.8) is satisfied, and L_e reduces to

$$L_e = -e_0^2/2a^2 \int \frac{\partial Z}{\partial x^i} \frac{\partial Z}{\partial x_i} dx^{(3)}. \quad (3.14)$$

Substituting for the intermediate solution

$$Z(\mathbf{r}) = \frac{1}{2}[(2\pi)^3 b \gamma]^{-1} \cdot \int \exp \{i \mathbf{p} \cdot \mathbf{r} / \hbar + [(\mathbf{p} \cdot \mathbf{v} / c)^2 - \mathbf{p}^2] / 4b^2\} / (\rho^2 - (\mathbf{p} \cdot \mathbf{v} / c)^2) \cdot d\mathbf{p}^{(3)} \quad (3.15)$$

one obtains

$$L_e = -e_0^2 / (2(2\pi)^{(3/2)} a \gamma). \quad (3.16)$$

This is the customary Lagrangian of a free particle with mass m_0 given by (2.17); the derived momentum and energy are $\frac{\partial L_e}{\partial \mathbf{v}} = m_0 \gamma \mathbf{v}$, and

$$-L_e + \mathbf{v} \cdot \frac{\partial L_e}{\partial \mathbf{v}} = m_0 \gamma c^2, \text{ a circumstance which justifies the adoption of } L_e$$

as the Lagrangian of the unperturbed electron. It will appear later, however, that m_0 is not the entire rest-mass of the particle, as a further contribution arises from the interaction with the electromagnetic field.

The remaining two terms L_f and L_{ef} in the Lagrangian, representing the electromagnetic field and the interaction respectively, are most conveniently expressed in terms of the pseudo-potentials $B_k = \exp(-P/4b^2) \cdot A_k$. Then L_f is given by Maxwell's formula

$$L_f = \frac{1}{2} \int \frac{\partial B^k}{\partial x_i} \frac{\partial B_k}{\partial x^i} dx^{(3)}, \quad (3.17)$$

and L_{ef} is reduced to

$$\begin{aligned} L_{ef} &= -e_0 v_k \int B^k(\mathbf{x}) e^{P/4b^2} \delta(\mathbf{r}) dx^{(3)} \\ &= -e_0 v_k (\sqrt{\pi} \cdot a)^{-3} \int B^k(\mathbf{x}) e^{-u^2/a^2} dx^{(3)} \end{aligned} \quad (3.18)$$

in the first approximation. This formula shows that the effect of the present theory is essentially the same as the introduction of an electron with a Gaussian charge distribution in the ordinary electrodynamics; this suggests an intuitive interpretation of the theory of reciprocity according to which measurements of the position of an electron are subject to an uncertainty of the order of the electronic radius, and the observed charge density is therefore not singular, but distributed in space according to the normal error law. The concept to the electron as a point singularity in the field A_k may nevertheless be retained.

Variation of the Lagrangian with respect to the co-ordinates of the electron will obviously give in this approximation an equation of motion which does not properly take account of the radiation damping. To obtain a second approximation, it is necessary to evaluate the expression $\exp P/4b^2 \cdot V_k \delta(\mathbf{x} - \mathbf{x}_0)$ for an electron with constant acceleration; this is

most easily effected by first writing it in the form

$$e^{P/4b^2} V_k \delta(\mathbf{x} - \mathbf{x}_s) = (2\pi\hbar)^{-3} \int \exp\left(-\frac{a^2}{4c^2} \frac{\partial^2}{\partial t^2}\right) \exp(i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}_s)/\hbar) d\mathbf{k}. \quad (3.19)$$

For brevity set $\sigma = -i\mathbf{k} \cdot \mathbf{x}_s/\hbar$, so that $d^3\sigma/dt^3$ vanishes by hypothesis. Then, since by straightforward computation

$$\frac{1}{(2n)!} \frac{\partial^{2n}}{\partial t^{2n}} \sigma^k = \sum_{m=0}^{k-n} \binom{k}{2m} \partial^{2m} \left(\frac{k-2m}{n-m} \right) \left(\frac{1}{2} \right) \partial^{2n-2m} \sigma^{k-n-m}, \quad (k \geq n) \quad (3.20)$$

one has

$$\exp\left(-\frac{a^2}{4c^2} \frac{\partial^2}{\partial t^2}\right) e^\sigma = e^\sigma \sum_{n=0}^{\infty} \left(-\frac{a^2}{4c^2}\right)^n \frac{(2n)!}{n!} \sum_{m=0}^n \frac{\partial^{2n} \left(\frac{1}{2}\partial\right)^{n-m}}{(2m)! (n-m)!}. \quad (3.21)$$

Putting $n = m + l$ and summing first over l and then over m , one obtains

$$\exp\left(-\frac{a^2}{4c^2} \frac{\partial^2}{\partial t^2}\right) e^\sigma = \zeta \exp\{-i\mathbf{k} \cdot \mathbf{x}/\hbar + \zeta^2 (\mathbf{k} \cdot \mathbf{v}/c)^2 / 4b^2\}, \quad (3.22)$$

where

$$\zeta = (1 - ia^2 \mathbf{k} \cdot \mathbf{x}_s / c^2 \hbar)^{-1/2}. \quad (3.23)$$

Similarly

$$\exp\left(-\frac{a^2}{4c^2} \frac{\partial^2}{\partial t^2}\right) \nabla e^\sigma = -\frac{a^2 \zeta^2}{c^2} \dot{\sigma} \nabla \exp\{-i\mathbf{k} \cdot \mathbf{x}_s/\hbar + \zeta^2 (\mathbf{k} \cdot \mathbf{v}/c)^2 / 4b^2\}. \quad (3.24)$$

Hence finally

$$\exp(P/4b^2) \cdot v_k \delta(\mathbf{x} - \mathbf{x}_s) = (2\pi\hbar)^{-3} \int \zeta \exp\{i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}_s)/\hbar - k^2/2b^2 + \zeta^2 (\mathbf{k} \cdot \mathbf{v}/c)^2 / 2b^2\} \\ (V_k + ia^2 \zeta^2 \mathbf{k} \cdot \dot{\mathbf{v}} v_k / c^2) d\mathbf{k}. \quad (3.25)$$

This result will be required in a later section. From it the exact solution of the field equations for a uniformly accelerated electron, valid at all distances from the singularity, can be written down at once as the usual solution of the equation $\square B_k = e_0 \exp(P/4b^2) v_k \delta(\mathbf{x} - \mathbf{x}_s)$. As the right-hand side is non-singular this solution is finite everywhere; but as it is rather too complicated to be of value in practical computations, an alternative method of solution which gives the values of the potentials for an arbitrary motion of the singularity at both large and small distances from the singularity has been developed, and is presented in the following section.

4. THE FIELD AND THE EQUATION OF MOTION OF A CHARGE IN ARBITRARY MOTION *

For an exact treatment of the field of a charge in arbitrary motion, the function (2.14) may be written in the form

$$L = L_0 + L_1.$$

* This section written by K. C. Cheng.

Here, according to (2.5) with (2.6) and (2.8), L_0 is

$$L_0 = -\frac{1}{2} \int \left(e^{-s^2/4} \frac{\partial A^k}{\partial x_i} \right) \left(e^{-s^2/4} \frac{\partial A^k}{\partial x^i} \right) dx^{(3)}, \quad (4.1)$$

and L_1 , represents the interaction term and self-energy in the form (2.17), thus:

$$L_1 = - \sum_i \int \left\{ \frac{e_i}{c} A^k \frac{dx_{(i)}^k}{ds} + m_0 \left(\frac{dx_{(i)}^k}{ds} \cdot \frac{dx_{(i)}^k}{ds} \right)^{1/2} \right\} \prod_{l=1}^4 \delta(x^l - x_{(i)}^l) dx^{(3)} ds, \quad (4.2)$$

where i distinguishes different electrons, as in § 2.

The total action is defined by

$$T = \int L dt, \quad (4.3)$$

from which the equation of motion of both field and particles can be derived from the principle of least action by variation with respect to A^k and the co-ordinates $x_{(i)}^k(s)$.

The variation with respect to A^k gives

$$\square e^{-s^2/4} A^k = - \sum_i \frac{e_i}{c} \int \frac{dx_{(i)}^k}{ds} \prod_{l=1}^4 \delta(x^l - x_{(i)}^l) ds. \quad (4.4)$$

On the other hand, variation with respect to $x_{(i)}^k(s)$ gives

$$\frac{d}{ds} \frac{\partial L}{\partial \frac{dx_{(i)}^k}{ds}} - \frac{\partial L}{\partial x_{(i)}^k} = 0,$$

or

$$\frac{d}{ds} \left\{ \frac{e_i}{c} A^k(x_{(i)}^k) + m_0 \frac{dx_{(i)}^k}{ds} \right\} = \frac{e_i}{c} \left(\frac{\partial A_1}{\partial x_{(i)}^k} \cdot \frac{dx_{(i)}^k}{ds} \right). \quad (4.5)$$

Setting

$$E_k^i = \frac{\partial A_k}{\partial x_i} - \frac{\partial A_i}{\partial x^k}, \quad f_{(i)k}^i = (E_k^i)_{x=x_{(i)}}$$

one finds the equation of motion of the charge

$$\frac{d^2}{ds^2} m_0 x_{(i)}^k = \frac{e_i}{c} f_{(i)k}^i \cdot \frac{dx_{(i)}^i}{ds}. \quad (4.6)$$

Writing for convenience

$$A_1(x_{(i)}) = A_{(i)1}^{(e)} + A_{(i)1}^{(o)}, \quad f_{(i)k}^i = f_{(i)k}^{(e)i} + f_{(i)k}^{(o)i},$$

where $A_{(ij)}^{(e)}$, $A_{(ij)}^{(e)}$ satisfy the equations

$$\begin{aligned} \square \exp(-\tfrac{1}{2}a^2 \square) A_{(ij)}^{(e)} &= -e_i/c \int \frac{d}{ds} x_{(i)} \prod_{k=1}^4 \delta(x^k - x_{(i)}^k) ds, \\ \square \exp(-\tfrac{1}{2}a^2 \square) A_{(ij)}^{(e)} &= -\sum_j' \frac{e_i}{c} \int \frac{d}{ds} x_{(i)} \prod_{k=1}^4 \delta(x^k - x_{(i)}^k) ds, \end{aligned} \quad (4.7)$$

and the dash sign excludes the term $i=j$, one has from (4.6)

$$m_0 \frac{d^2 x_{(i)}^k}{ds^2} - \frac{e_i}{c} f_{(i)}^{(e)} \frac{dx_{(i)}^j}{ds} = \frac{e_i}{c} f_{(i)}^{(e)} \frac{dx_{(i)}^j}{ds}. \quad (4.8)$$

The second term on the left-hand side of (4.8) will be worked out and discussed below.

According to (4.4) the field equation can be written in the following form:—

$$\square \exp(-\tfrac{1}{2}a^2 \square) \cdot A^k = -e_0 \int \delta(\mathbf{x} - \mathbf{x}_0(t_s)) \delta(t - t_s) v^k / c dt_s, \quad (4.9)$$

where v^k is the four vector (\mathbf{v}, c) representing the velocity of the electron, and the proper time s has been replaced by the time co-ordinate $t_s = \frac{1}{c} x_{s4}$.

From now on, for the convenience in discussion, attention will be confined to the problem of one electron. Writing $D = \frac{1}{c} \frac{\partial}{\partial t}$ and $\Delta = \frac{\partial}{\partial \mathbf{x}} \cdot \frac{\partial}{\partial \mathbf{x}}$, so that $\square = \Delta - D^2$, and operating on both sides of (4.9) by $\exp(\tfrac{1}{2}a^2 \Delta)$, one has

$$\square \exp(\tfrac{1}{2}a^2 D^2) \cdot A^k = -e_0 \int \exp(\tfrac{1}{2}a^2 \Delta) \delta(\mathbf{x} - \mathbf{x}_s) \delta(t - t_s) v^k / c dt_s. \quad (4.10)$$

Using the relation

$$\exp(\tfrac{1}{2}a^2 \Delta) \cdot \delta(\mathbf{x} - \mathbf{x}_s) = (2\pi)^{-3/2} a^{-3} \exp(-(\mathbf{x} - \mathbf{x}_s)^2 / 2a^2),$$

which can be easily verified by comparing the Fourier components of both sides, (4.10) becomes

$$\square \exp(\tfrac{1}{2}a^2 D^2) A^k = -e_0 (\sqrt{(2\pi) \cdot a})^{-3} c^{-1} \int \delta(t - t_s) v^k \exp\{-[\mathbf{x} - \mathbf{x}_s(t_s)]^2 / 2a^2\} dt_s. \quad (4.11)$$

To preserve the customary idea of causality one has to choose, among the possible solutions of the D'Alembertian differential equation (4.11), the retarded one, namely

$$\begin{aligned} A^k &= e_0 (2a^3 c)^{-1} (2\pi)^{-5/2} \exp(-\tfrac{1}{2}a^2 D^2) \int v^k(t_s) \delta(t - t_s - |\mathbf{x} - \mathbf{x}'|/c) |\mathbf{x} - \mathbf{x}'|^{-1} \\ &\quad \exp[-\{\mathbf{x}' - \mathbf{x}_s(t_s)\}^2 / 2a^2] dt_s dx_s^{(3)}, \end{aligned} \quad (4.12)$$

where the third term inside the δ -function takes account of the retardation.

Introducing the abbreviations

$$\mathbf{R} = \mathbf{x}' - \mathbf{x}, \quad R = |\mathbf{R}|, \quad \mathbf{r} = \mathbf{x} - \mathbf{x}_s, \quad r = |\mathbf{r}|, \quad \cos \theta = \mathbf{R} \cdot \mathbf{r} / Rr,$$

one has

$$A^k = \frac{1}{2} \epsilon_0 (a^2 c)^{-1} (2\pi)^{-5/2} \exp(-\frac{1}{2} a^2 D^2) \int R v^k \delta(t - t_s - R/c) \sin \theta \cdot \exp[-(R^2 + r^2 + 2rR \cos \theta)/2a^2] d\theta d\phi dR dt_s. \quad (4.13)$$

It should be noted that for $t_s > t$, $\delta(t - t_s - R/c) = 0$; thus the region of integration for t_s is between the limits t and $-\infty$. Performing the integration first over ϕ and R , then over θ , one obtains

$$\begin{aligned} A^k &= \frac{1}{2} \epsilon_0 c (\sqrt{(2\pi) \cdot a})^{-3} \exp(-\frac{1}{2} a^2 D^2) \int_{-\infty}^t dt_s \int_{-\pi/2}^{\pi/2} v^k(t - t_s) \sin \theta \\ &\quad \exp[-\{c^2(t - t_s)^2 + 2rc(t - t_s) \cos \theta + r^2\}/2a^2] d\theta \\ &= \frac{1}{2} \epsilon_0 c (\sqrt{(2\pi) \cdot a})^{-1} \exp(-\frac{1}{2} a^2 D^2) \int_{-\infty}^t a^2 v^k(r) c^{-1} \{ \exp[-(c(t - t_s) - r)^2/2a^2] \\ &\quad - \exp[-(c(t - t_s) + r)^2/2a^2] \} dt_s. \end{aligned} \quad (4.14)$$

Let

$$\zeta = r(t_s) - c(t - t_s), \quad \eta = -r(t_s) - c(t - t_s), \quad (4.15)$$

then

$$\frac{d\zeta}{dt_s} = c - v_r, \quad \frac{d\eta}{dt_s} = c + v_r, \quad (4.16)$$

where $v_r = (\mathbf{v} \cdot \mathbf{r})/r$.

Now any function $f(\mathbf{x}, t_s)$ of the variables \mathbf{x} and t_s can be expressed in terms of \mathbf{x} and ζ or η with the help of (4.15); then it is obvious that

$$\frac{\partial}{\partial \zeta} f\{\mathbf{x}, t_s[\mathbf{x}, (\zeta + ct)]\} = \frac{1}{c} \frac{\partial}{\partial t} f\{\mathbf{x}, t_s[\mathbf{x}, (\zeta + ct)]\}, \quad (4.17)$$

and, by Taylor's expansion,

$$f\{\mathbf{x}, t_s[\mathbf{x}, (\zeta + ct)]\} = \sum_0^{\infty} \frac{\zeta^n}{n!} D^n f\{\mathbf{x}, t_s[\mathbf{x}, (o + ct)]\}, \quad (4.18)$$

where, in view of (4.15),

$$f\{\mathbf{x}, t_s[\mathbf{x}, (o + ct)]\} = [f(\mathbf{x}, t_s)]_{\text{ret.}},$$

with t_s defined by the condition $r(t_s) = c(t - t_s)$, and "ret." outside the bracket denotes as usual the retarded value.

Similarly, for any function $g(\mathbf{x}, t_s)$, one has

$$g(\mathbf{x}, t_s) = \sum_0^{\infty} \frac{\eta^n}{n!} D^n [g(\mathbf{x}, t_s)]_{\text{adv.}}, \quad (4.19)$$

where t_0 is determined by the condition $r(t_0) = c(t_0 - t)$, and "adv." denotes the advanced value.

With the help of the definitions (4.15) and (4.16), (4.14) becomes

$$A^k = \frac{e_0}{2(2\pi)^{3/2}a} e^{-a^2 D^2/2} \left\{ \int_{-\infty}^{r(t)} \frac{e^{-\zeta^2/2a^2}}{r(\zeta)(c-v_r)} v^k d\zeta - \int_{-\infty}^{-r(t)} \frac{e^{-\eta^2/2a^2}}{r(\eta)(c+v_r)} v^k d\eta \right\}. \quad (4.20)$$

The two cases where $r(t)$ is large or small compared with a will be discussed separately below in (i) and (ii). In (iii) the explicit equation of motion of a charge will be deduced and be discussed.

(i) *For Large Distances $r \gg a$*

The second term in (4.20) is negligible because of the smallness of the exponential factor, and, as r/a is large, one can replace the limits $r(t)$ by infinity. Then

$$A^k = \frac{e_0}{2(2\pi)^{3/2}a} e^{-a^2 D^2/2} \int_{-\infty}^{\infty} \frac{e^{-\zeta^2/2a^2}}{r(\zeta)(c-v_r)} v^k d\zeta. \quad (4.21)$$

On using equation (4.18) one obtains

$$\frac{v^k}{r(\zeta)(c-v_r)} = \sum_0^{\infty} \frac{1}{n!} \zeta^n D^n \left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}}$$

Hence

$$\begin{aligned} A^k &= \frac{e_0}{(2\pi)^{3/2}a} e^{-a^2 D^2/2} \sum_0^{\infty} \int_{-\infty}^{\infty} \frac{\zeta^n}{n!} e^{-\zeta^2/2a^2} d\zeta D^n \left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}} \\ &= \frac{e_0}{4\pi} e^{-a^2 D^2/2} \sum_0^{\infty} \frac{1}{n!} \left(\frac{a^2 D^2}{2} \right)^n \left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}} \\ &= \frac{e_0}{4\pi} \left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}}. \end{aligned} \quad (4.22)$$

(4.22) shows that at large distances the four-potentials are identical with those obtained by Wiechert. The error introduced in replacing $r(t)$ by infinity is of the order of $e^{-r^2/2a^2}$, which is very small when r/a is large

(ii) *For Small Distances $r \ll a$*

Writing

$$A^k = \frac{e_0}{2(2\pi)^{3/2}a} (P^k + Q^k),$$

where

$$P^k = e^{-a^2 D^2/2} \left\{ \int_{-\infty}^0 \frac{e^{-\zeta^2/2a^2}}{r(\zeta-v_r)} v^k d\zeta - \int_{-\infty}^0 \frac{e^{-\eta^2/2a^2}}{r(\zeta+v_r)} v^k d\eta \right\},$$

$$Q^k = e^{-a^2 D^2/2} \left\{ \int_0^{r(0)} \frac{e^{-\zeta^2/2a^2}}{r(\zeta-v_r)} v^k d\zeta + \int_0^{-r(0)} \frac{e^{-\eta^2/2a^2}}{r(\zeta+v_r)} v^k d\eta \right\}. \quad (4.23)$$

Using (4.18) one obtains

$$\int_{-\infty}^0 \frac{e^{-\zeta^2/2a^2}}{r(\zeta)(\zeta-v_r)} v^k d\zeta = \sum_0^{\infty} (-)^n \int_0^{\infty} \frac{1}{n!} \zeta^n e^{-\zeta^2/2a^2} d\zeta D^n \left[\frac{v^k}{r(\zeta-v_r)} \right]_{\text{ret.}}$$

$$= \frac{1}{2} \sum_0^{\infty} \frac{(-)^n}{n!} (\sqrt{2a})^{n+1} \Gamma\left(\frac{n+1}{2}\right) D^n \left[\frac{v^k}{r(\zeta-v_r)} \right]_{\text{ret.}};$$

and by separating the even and odd powers of D one has

$$\left(\frac{1}{2}\pi\right)^{\frac{1}{2}} a e^{a^2 D^2/2} \left[\frac{v^k}{r(\zeta-v_r)} \right]_{\text{ret.}} - \frac{1}{2} \sum_0^{\infty} \frac{(\sqrt{2a})^{2n+1} n!}{(2n+1)!} D^{2n+1} \left[\frac{v^k}{r(\zeta-v_r)} \right]_{\text{ret.}}$$

By a similar procedure one obtains

$$\int_{-\infty}^0 \frac{e^{-\eta^2/2a^2}}{r(\eta)(\zeta+v_r)} v^k d\eta = \left(\frac{1}{2}\pi\right)^{\frac{1}{2}} a e^{a^2 D^2/2} \left[\frac{v^k}{r(\zeta+v_r)} \right]_{\text{adv.}} - \frac{1}{2} \sum_0^{\infty} \frac{(\sqrt{2a})^{2n+1} n!}{(2n+1)!} D^{2n+1} \left[\frac{v^k}{r(\zeta+v_r)} \right]_{\text{adv.}}$$

Hence

$$P^k = a \left(\frac{1}{2}\pi\right)^{\frac{1}{2}} \left\{ \left[\frac{v^k}{r(\zeta-v_r)} \right]_{\text{ret.}} - \left[\frac{v^k}{r(\zeta+v_r)} \right]_{\text{adv.}} \right\} + R^k$$

with

$$R^k = -\frac{1}{2} e^{-a^2 D^2/2} \sum_0^{\infty} \frac{D^{2n+1} n!}{(2n+1)!} (\sqrt{2a})^{2n+1} \left\{ \left[\frac{v^k}{r(\zeta-v_r)} \right]_{\text{ret.}} - \left[\frac{v^k}{r(\zeta+v_r)} \right]_{\text{adv.}} \right\}. \quad (4.24)$$

Expanding R^k in powers of a one has

$$R^k = -2 \sum_1^{\infty} \left(\frac{1}{2}a^2\right)^s (-)^{s+1} \left\{ \sum_0^{s-1} \frac{(-)^m 2^{2m} m!}{(s-m-1)!(2m+1)!} \right\} D^{2s-1} \left\{ \left[\frac{v^k}{r(\zeta-v_r)} \right]_{\text{ret.}} - \left[\frac{v^k}{r(\zeta+v_r)} \right]_{\text{adv.}} \right\}.$$

It is shown in Appendix I that

$$\sum_0^{s-1} \frac{2^{2m} (-)^m m!}{(s-m-1)!(2m+1)!} = \frac{1}{(2s-1)(s-1)!}.$$

Then

$$R^k = -\sum_{s=1}^{\infty} \left(\frac{1}{2}a^2\right)^s (-)^s \frac{1}{\Gamma(s)(s-\frac{1}{2})} D^{2s-1} \left\{ \left[\frac{v^k}{r(\zeta-v_r)} \right]_{\text{ret.}} - \left[\frac{v^k}{r(\zeta+v_r)} \right]_{\text{adv.}} \right\}.$$

It is well known that the solutions $\left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}}$, $\left[\frac{v^k}{r(c+v_r)} \right]_{\text{adv.}}$ can be expanded in power series of $D^n(r^{n-1}v^k)$; * thus

$$\begin{aligned} \left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}} &= \sum_0^{\infty} \frac{(-)^n}{n!} D^n[r^{n-1}v^k/c], \\ \left[\frac{v^k}{r(c+v_r)} \right]_{\text{adv.}} &= \sum_0^{\infty} \frac{1}{n!} D^n[r^{n-1}v^k/c], \end{aligned} \quad (4.25)$$

and hence

$$\left\{ \left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}} - \left[\frac{v^k}{r(c+v_r)} \right]_{\text{adv.}} \right\} = -2 \sum_{\mu=0}^{\infty} \frac{D^{2\mu+1}}{(2\mu+1)!} r^{2\mu} v^k/c; \quad (4.26)$$

also

$$R^k = \sum_{\mu, s=0}^{\infty} \left(\frac{1}{2} a^2 \right)^s D^{2s+2\mu} \frac{4(-)^s}{\Gamma(s)\Gamma(2\mu+2)(1-2s)} r^{2\mu} v^k/c. \quad (4.27)$$

Now for the term Q^k one has, by applying (4.18) and (4.19),

$$\begin{aligned} Q^k &= e^{-a^2 D^2/2} \sum_0^{\infty} \frac{1}{n!} \left\{ \int_0^{r(t)} e^{-\zeta^2/2a^2} \zeta^n d\zeta D^n \left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}} \right. \\ &\quad \left. + (-)^n \int_0^{r(t)} e^{-\eta^2/2a^2} \eta^n d\eta D^n \left[\frac{v^k}{r(c+v_r)} \right]_{\text{adv.}} \right\} \end{aligned} \quad (4.28)$$

By using (4.25) one obtains

$$\begin{aligned} &\sum_0^{\infty} \frac{1}{n!} \zeta^n D^n \left\{ \left[\frac{v^k}{r(c-v_r)} \right]_{\text{ret.}} + (-)^n \left[\frac{v^k}{r(c+v_r)} \right]_{\text{adv.}} \right\} \\ &= 2 \sum_{n=0, m=0}^{\infty} \left\{ \frac{\zeta^{2n-2m} D^{2n}}{(2m)!(2n-2m)!} r^{2m-1} v^k/c - \frac{\zeta^{2n-2m+1} D^{2n+2}}{(2m+1)!(2n-2m+1)!} r^{2m} v^k/c \right\} \\ &= 2 \sum_0^{\infty} \frac{1}{(2n)!} D^{2n} \frac{1}{r} (\zeta-r)^{2n} v^k/c. \end{aligned}$$

Since

$$D^{2n} \int_0^{r(t)} (\zeta-r)^{2n} v^k/c d\zeta = \int_0^{r(t)} D^{2n} (\zeta-r)^{2n} v^k/c d\zeta,$$

because

$$r(t) D^{2n-1} (\zeta-r)_{\zeta=r}^{2n} = 0,$$

* Cf. Frenkel, *Elektrodynamik*, 1, 184 (1926).

one has

$$Q^k = 2e^{-a^2 D^2/2} \sum_{n=0}^{\infty} \frac{1}{(2n)!} D^{2n} \frac{1}{r} \int_0^r (\zeta - r)^{2n} e^{-\zeta^2/2a^2} d\zeta v^k/c$$

$$= 2e^{-a^2 D^2/2} \sum_{n,m=0}^{\infty} D^{2n} \left[\frac{(2m)! (-)^m}{m! (2m+2n+1)!} \left(\frac{1}{2}a^2\right)^m r^{2m+2n} v^k/c \right].$$

Expanding Q^k in powers of a and keeping the first power in r , but all powers of \dot{r} , \ddot{r} , etc., then one has

$$Q^k = 2 \sum_{n=0}^{\infty} a^{2n} Q_n^k, \quad (4.29)$$

where

$$Q_n^k = (-)^n \left(\frac{1}{2}\right)^n \sum_{\mu=0}^{\infty} \frac{1}{(2\mu+1)!} \left\{ \sum_{m=0}^{\mu} \frac{(2m)!}{m! (s+m)!} \left(\frac{1}{2}\right)^{2m} \right\} D^{2s+2\mu} r^{2\mu} v^k/c. \quad (4.30)$$

It is shown in Appendix II that

$$\sum_{m=0}^{\mu} \frac{(2m)!}{m! (s+m)!} \left(\frac{1}{2}\right)^{2m} = \frac{2}{\Gamma(\frac{1}{2})(1-2s)} \left\{ \frac{\Gamma(\mu+\frac{3}{2})}{\Gamma(s+\mu+1)} - \frac{\Gamma(\frac{1}{2})}{\Gamma(s)} \right\}. \quad (4.31)$$

Thus, combining (4.24), (4.27) and (4.30), and using (4.31), one obtains

$$A^k = \frac{e_0}{2(2\pi)^{3/2}a} (P^k + Q^k)$$

$$= \frac{e_0}{8\pi} \left\{ \left[\frac{v^k}{(c-v_r)r} \right]_{\text{ret.}} - \left[\frac{v^k}{r(c+v_r)} \right]_{\text{adv.}} \right\} + \frac{e_0}{(2\pi)^{3/2}a} \sum_{n=0}^{\infty} a^{2n} U_n^k, \quad (4.32)$$

where

$$U_n^k = (-)^n \left(\frac{1}{2}\right)^{n-1} \frac{1}{\Gamma(s)(1-2s)} \sum_{\mu=0}^{\infty} \frac{\Gamma(\mu+\frac{3}{2})}{(2\mu+1)!} \frac{D^{2\mu+2s}}{(\mu+s)!} r^{2\mu} v^k/c. \quad (4.33)$$

For the purpose of illustration the first two U_n^k 's are computed in Appendix III. The following are the results of computation:—

$$U_0^k = \frac{1}{c} \gamma v^k - \frac{1}{c^3} \left\{ \gamma^3 (\nabla \cdot \mathbf{r}) \dot{v}^k + \frac{1}{2} \gamma^3 (\dot{\nabla} \cdot \mathbf{r}) v^k + \frac{3}{2c^2} \gamma^3 (\dot{\nabla} \cdot \nabla) ((\nabla \cdot \mathbf{r}) v^k) \right\}, \quad (4.34)$$

$$2U_1^k = \frac{1}{c^2} \gamma^3 \{ (\ddot{\nabla} \cdot \nabla) v^k + \frac{3}{2} (\dot{\nabla} \cdot \dot{\nabla}) v^k + 3 (\dot{\nabla} \cdot \nabla) \dot{v}^k \} + \frac{1}{c^2} \gamma^3 \ddot{v}^k$$

$$+ \frac{15}{4c^2} (\dot{\nabla} \cdot \nabla)^2 v^k - \frac{1}{c^4} \left\{ \gamma^3 (\nabla \cdot \mathbf{r}) \ddot{v}^k + \left[\frac{3}{2} \gamma^3 (\dot{\nabla} \cdot \mathbf{r}) + \frac{15}{2c^2} \gamma^3 (\dot{\nabla} \cdot \nabla) (\nabla \cdot \mathbf{r}) \right] \dot{v}^k \right.$$

$$\left. + \left[\gamma^3 (\ddot{\nabla} \cdot \mathbf{r}) + \frac{15}{2c^2} \gamma^3 (\ddot{\nabla} \cdot \nabla) (\dot{\nabla} \cdot \mathbf{r}) + \frac{5\gamma^3}{c^2} (\ddot{\nabla} \cdot \nabla) (\nabla \cdot \mathbf{r}) + \frac{15}{2c^2} \gamma^3 (\nabla \cdot \dot{\nabla}) (\nabla \cdot \mathbf{r}) \right] v^k \right\}$$

$$\begin{aligned}
& + \left[\frac{1}{2} \gamma^3 (\ddot{\mathbf{v}} \cdot \mathbf{r}) + \frac{5\gamma^7}{2c^2} (\dot{\mathbf{v}} \cdot \mathbf{v})(\ddot{\mathbf{v}} \cdot \mathbf{r}) + \frac{5\gamma^7}{2c^2} (\ddot{\mathbf{v}} \cdot \mathbf{v})(\dot{\mathbf{v}} \cdot \mathbf{r}) + \frac{15\gamma^7}{8c^2} (\dot{\mathbf{v}} \cdot \dot{\mathbf{v}})(\ddot{\mathbf{v}} \cdot \mathbf{r}) \right. \\
& + \frac{105}{8c^4} \gamma^3 (\dot{\mathbf{v}} \cdot \mathbf{v})^2 (\dot{\mathbf{v}} \cdot \mathbf{r}) + \frac{5}{4c^2} \gamma^7 (\ddot{\mathbf{v}} \cdot \mathbf{v})(\mathbf{v} \cdot \mathbf{r}) + \frac{5\gamma^7}{2c^2} (\dot{\mathbf{v}} \cdot \dot{\mathbf{v}})(\mathbf{v} \cdot \mathbf{r}) + \frac{35}{2c^4} \gamma^3 (\dot{\mathbf{v}} \cdot \mathbf{v})(\ddot{\mathbf{v}} \cdot \mathbf{v})(\mathbf{v} \cdot \mathbf{r}) \\
& \left. + \frac{105}{8c^4} \gamma^3 (\dot{\mathbf{v}} \cdot \dot{\mathbf{v}})(\dot{\mathbf{v}} \cdot \mathbf{v})(\mathbf{v} \cdot \mathbf{r}) + \frac{315}{8c^6} \gamma^{11} (\dot{\mathbf{v}} \cdot \mathbf{v})^2 (\mathbf{v} \cdot \mathbf{r}) \right] v^k, \quad (4.35)
\end{aligned}$$

where $\dot{\mathbf{v}} = \frac{d}{dt} \mathbf{v}$, etc., and $\gamma = \frac{1}{\sqrt{1 - \mathbf{v}^2/c^2}}$.

It follows from (4.32) and (4.33) that the potentials are all finite at the origin.

(iii) The Equation of Motion of an Electron

According to (4.8) the equation of motion of an electron is the following:—

$$\frac{d^2}{ds^2} (m_0 x_s^\mu) - e_0/c f_{\mu}^{(\nu)} \frac{d}{ds} x_s^\mu = \frac{e_0}{c} f_{\mu}^{(\nu)} \frac{d}{ds} x_s^\mu, \quad (4.36)$$

where

$$f_{\mu}^{(\nu)} = \left(\frac{\partial}{\partial x_\nu} A^{(\nu)\mu} - \frac{\partial}{\partial x_\mu} A^{(\nu)\nu} \right)_{x=x_0}.$$

Using the three-dimensional space description, one has for $\nu = 1, 2, 3$ the following vectorial equation. With $A_4 = \phi c$, $\mathbf{A} = (A_1, A_2, A_3)$,

$$m_0 \frac{d}{dt} \gamma \mathbf{v} - e_0 \left(\mathbf{E}^{(\nu)} + \frac{1}{c} \mathbf{v} \wedge \mathbf{H}^{(\nu)} \right) = e_0 \left(\mathbf{E}^{(\nu)} + \frac{1}{c} \mathbf{v} \wedge \mathbf{H}^{(\nu)} \right), \quad (4.37)$$

where

$$\mathbf{E} = -\frac{\partial}{\partial \mathbf{x}} \phi - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{A}, \quad \mathbf{H} = \text{curl } \mathbf{A}.$$

Writing $\mathbf{A}^{(\nu)}$, $\phi^{(\nu)}$, etc. in powers of a ,

$$\mathbf{A}^{(\nu)} = \sum_{m=-1}^{\infty} a^m \mathbf{A}_m^{(\nu)}, \quad \phi^{(\nu)} = \sum_{m=-1}^{\infty} a^m \phi_m^{(\nu)}, \quad \text{etc.,}$$

then, according to (4.32) and (4.34) one obtains

$$\begin{aligned}
\mathbf{E}_{-1}^{(\nu)} &= -\frac{\partial}{\partial \mathbf{x}} \phi_{-1}^{(\nu)} - \frac{1}{c} \frac{\partial}{\partial t} \mathbf{A}_{-1}^{(\nu)} \\
&= e_0 (ac)^{-1} (2\pi)^{-1/2} \left\{ -\frac{d}{dt} \gamma \mathbf{v} + \gamma^3/c^2 (\dot{\mathbf{v}} \cdot \mathbf{v}) \mathbf{v} + \gamma^3 \left(\frac{1}{2} - \mathbf{v}^2/c^2 \right) \dot{\mathbf{v}} \right\}, \quad (4.38)
\end{aligned}$$

and similarly

$$\begin{aligned} \frac{1}{c} \nabla \wedge \mathbf{H}_{-1}^{(e)} &= \frac{1}{c} \nabla \wedge \frac{\partial}{\partial \mathbf{x}} \wedge \mathbf{A}_{-1}^{(e)} \\ &= e_0 (ac^2)^{-1} (2\pi)^{-3/2} \left\{ \frac{1}{2} c^{-2} \gamma^2 v^2 \dot{\mathbf{v}} - \frac{1}{2} c^{-2} \gamma^2 (\dot{\mathbf{v}} \cdot \nabla) \mathbf{v} \right\}. \end{aligned} \quad (4.39)$$

The addition of the two terms given by (4.39) and (4.38) gives the surprisingly simple result

$$-e_0 \left(\mathbf{E}_{-1} + \frac{1}{c} \nabla \wedge \mathbf{H}_{-1} \right) = -\frac{1}{2} e_0^2 (ac^2)^{-1} (2\pi)^{-3/2} \gamma^{-1} \frac{d^2}{ds^2} \mathbf{x}.$$

Now the term

$$\begin{aligned} -e_0 \left(\mathbf{E}_0 + \frac{1}{c} \nabla \wedge \mathbf{H}_0 \right) &= e_0^2 \gamma^{-1} \frac{dx^i}{ds} \left[\frac{\partial}{\partial x_i} \{ [\nabla/r(c-v_r)]_{\text{ret.}} - [\nabla/r(c+v_r)]_{\text{adv.}} \} \right. \\ &\quad \left. - \frac{\partial}{\partial \mathbf{x}} \{ [v_i/r(c-v_r)]_{\text{ret.}} - [v_i/r(c+v_r)]_{\text{adv.}} \} \right] \end{aligned}$$

is simply Dirac's radiation damping term, which, being evaluated, is equal to

$$-\frac{e_0^2}{6\pi c^2 \gamma} \left\{ \frac{d^2}{ds^2} \mathbf{x} + c^{-2} \left(\frac{d^2 x^k}{ds^2} \cdot \frac{d^2 x_k}{ds^2} \right) \frac{d}{ds} \mathbf{x} \right\}^*.$$

Thus, by neglecting the other powers of a , one obtains exactly Dirac's equation of motion of an electron:

$$\begin{aligned} \{ m_0 + \frac{1}{2} e_0^2 (ac^2)^{-1} (2\pi)^{-3/2} \} \frac{d^2}{ds^2} \mathbf{x} &- e_0^2 / 6\pi c^2 \cdot \left\{ \frac{d^2}{ds^2} \mathbf{x} + c^{-2} \left(\frac{d^2 x^k}{ds^2} \cdot \frac{d^2 x_k}{ds^2} \right) \frac{d\mathbf{x}}{ds} \right\} \\ &= e_0 \frac{d\mathbf{t}}{ds} \cdot (\mathbf{E}^{(e)} + c^{-1} \nabla \wedge \mathbf{H}^{(e)}), \end{aligned} \quad (4.40)$$

which is obviously relativistically invariant.

However, there is a difference between (4.40) and Dirac's equation of motion if one has under consideration several electrons interacting with one another at distances smaller or comparable with the fundamental length a . The difference arises from the right-hand side of the equation of motion. The external field due to other sources $e_0(\mathbf{E}^{(e)} + c^{-1} \nabla \wedge \mathbf{H}^{(e)})$ consists, according to Dirac's theory, only of the retarded potentials. In the present theory this conclusion is correct if the particles are separated by distances greater than a ; for at large distances, the field acting on each charge is shown in (4.32) to be identical with that derived from the retarded potentials, whereas, if r is smaller or comparable with a , this conclusion no longer holds: the potentials given by (4.22) have to be replaced by

* This term was actually given many years earlier by W. Behrens and E. Hecke, *Nachrichten der K. Gesellschaft der Wissenschaften zu Göttingen* (1912).

(4.32), in which retarded and advanced potentials play symmetrical rôles.

The first correction to the equation (4.40) proportional to a will now be evaluated. By applying the same procedure as that by which (4.40) was derived, the following equation may be obtained:—

$$\left. \begin{aligned} & 2m_0 \frac{d^2}{ds^2} \mathbf{x} - \frac{e_0^2}{6\pi c^3} \left\{ \frac{d^2 \mathbf{x}}{ds^2} + c^{-1} \left(\frac{d^2 x^k}{ds^2} \cdot \frac{dx_k}{ds} \right) \frac{d\mathbf{x}}{ds} \right\} \\ & + \frac{1}{2} e_0^2 a c^{-4} (2\pi)^{-3/2} \left\{ \frac{3}{4} \frac{d^4}{ds^4} \mathbf{x} + 9(8c^2)^{-1} \frac{d}{ds} \left[\left(\frac{d^2 x^k}{ds^2} \cdot \frac{d^2 x_k}{ds^2} \right) \frac{d}{ds} \mathbf{x} \right] \right\} \\ & - e_0 \frac{dt}{ds} (\mathbf{E}^{(a)} + \nabla \wedge \mathbf{H}^{(a)}/c), \end{aligned} \right\} \quad (4.41)$$

where

$$m_0 = \frac{1}{2} e_0^2 (ac^2)^{-1} (2\pi)^{-3/2}.$$

The detailed computational procedure is in Appendix III. This correction corresponds to that for an electron of finite radius which has been in pre-relativistic times studied by several authors for a rigid model (Herglotz, 1903; Sommerfeld, 1904; Krazér, 1905; Born, 1909). If the relativistic terms are neglected, the new formulæ are almost identical with the old ones. The present form has the advantage of being relativistically invariant in every power of a . It shows that a relativistically invariant theory of an electron with a structure is possible and that only the meaning of the radius has to be re-interpreted.

5. THE HAMILTONIAN FORMULATION *

The Hamiltonian energy and total momentum for an electron in interaction with an electromagnetic field are readily derived from the Lagrangian defined by (2.18). On account of the auxiliary condition (2.8), the Lagrangian may be written in the simplified form

$$\left. \begin{aligned} L &= \frac{1}{2} \int \langle F A^k A'_k \rangle dx^{(3)} - \frac{e_0}{c} v_k A^k(\mathbf{x}_e) - m_0 c^2 (1 - v^2/c^2)^{1/2}, \\ F &= F(p, p') = -\hbar^{-1} e^{-(P+P')/4\hbar} p'_m p^m. \end{aligned} \right\} \quad (5.1)$$

In order to derive the energy and momentum, the general method which was explained in Q.T.R. may be applied to the first term, and the usual

* Following sections written by H. S. Green.

methods suffice for the remainder. The Hamiltonian energy so obtained is

$$\epsilon P_4 = \frac{1}{2} \int \langle (-F + G^k p_k + G^k p'_k) A^k A'_k \rangle dx^{(3)} + e_0 A_4(\mathbf{x}_e) + m_0 c^2 (1 - v^2/c^2)^{-1/2}, \quad (5.2)$$

where G^i and G^i' are defined so as to satisfy

$$\begin{aligned} F(p', p') &= F(p, p') + (p'_1 - p_1) G^1(p, p'), \\ F(p, p) &= F(p, p') + (p_1 - p'_1) G^{1'}(p, p'), \end{aligned} \quad (5.3)$$

and are explicitly

$$\begin{aligned} G_i &= -\hbar^{-2} e^{-(P+P')/4\omega^2} \left\{ p'_i + (p_i + p'_i) \left(\frac{e^{(P-P')/4\omega^2} - 1}{P' - P} \right) P' \right\}, \\ G_i' &= -\hbar^{-2} e^{-(P+P')/4\omega^2} \left\{ p_i + (p_i + p'_i) \left(\frac{e^{(P'-P)/4\omega^2} - 1}{P - P'} \right) P \right\}. \end{aligned} \quad (5.4)$$

The total momentum is defined by

$$\mathbf{P} = \frac{1}{2c} \int \langle (G^k \mathbf{p} + G^{k'} \mathbf{p}') A^k A'_k \rangle dx^{(3)} + \frac{e_0}{c} \mathbf{A}(\mathbf{x}_e) + m_0 \mathbf{v} (1 - v^2/c^2)^{-1/2}, \quad (5.5)$$

so as to form, together with P_4 , a four-vector.

It can be shown that, by virtue of the field equations and equation of motion of the electron,

$$\left. \begin{aligned} F(p, p) A_i(x) &= \frac{e_0}{c} v_i \delta(\mathbf{x} - \mathbf{x}_e), \\ \frac{d}{dt} \left\{ m v_i (1 - v^2/c^2)^{-1/2} + \frac{e_0}{c} A_i(\mathbf{x}_e) \right\} &= \frac{e_0}{c} v_k \frac{\partial A^k}{\partial x^i}(\mathbf{x}_e), \end{aligned} \right\} \quad (5.6)$$

the conservation law $dP_i/dt = 0$ is satisfied. For, from the first of the equations (5.6) it follows that

$$\int \langle \{ F(p', p') p'_i - F(p, p) p_i \} A^k A'_k \rangle dx^{(3)} = \frac{2i\hbar e_0 v^k}{c} \frac{\partial A_k}{\partial x^i}(\mathbf{x}_{(t)}), \quad (5.7)$$

and the left-hand side of this equation is, according to (5.3),

$$\int \langle \{ -(p'_i - p_i) F + (p'_k - p_k) (G^k p_i + G^{k'} p'_i) \} A^k A'_k \rangle dx^{(3)}.$$

Since, if $J(p, p')$ is any function of p_k and p'_k ,

$$\begin{aligned} \int \langle (p'_i - p_i) J A^k A'_k \rangle dx^{(3)} &= -i\hbar \int \frac{\partial}{\partial x^i} \langle J A^k A'_k \rangle dx^{(3)} \\ &= -\frac{i\hbar}{c} \delta_i^j \frac{d}{dt} \int \langle J A^k A'_k \rangle dx^{(3)}, \end{aligned} \quad (5.8)$$

the equation (5.7) reduces to

$$\frac{1}{2} \frac{d}{dt} \int \langle (-F\delta_1^4 + G^4 p_1 + G^4 p'_1) A^k A'_k \rangle dx^{(3)} = -c_0 v^k \frac{\partial A_k}{\partial x^i}(\mathbf{x}_{(t)}). \quad (5.9)$$

Combining (5.9) with the second of equations (5.6), the desired result $dP_i/dt=0$ is immediate.

The Hamiltonian cP_4 may be expressed in an alternative form by substituting in (5.2) the expression

$$F = F(p, p') = \frac{1}{2}\{F(p', p') + F(p, p)\} + \frac{1}{2}(p'_1 - p_1)(G' - G) \quad (5.10)$$

derived from (5.3). The term arising from $\frac{1}{2}\{F(p', p') + F(p, p)\}$ can be simplified with the help of the field equations (5.6), and it can be seen from (5.8) that the term containing $(\mathbf{p}' - \mathbf{p}) \cdot (\mathbf{G}' - \mathbf{G})$ will contribute nothing; the result is therefore

$$cP_4 = \frac{1}{2} \int \langle (p_4 + p'_4)(G_4 + G'_4) A^k A'_k \rangle dx^{(3)} + \frac{e_0}{c} \mathbf{v} \cdot \mathbf{A}(\mathbf{x}_{(t)}) + m_0 c^2 (1 - v^2/c^2)^{-\frac{1}{2}} \quad (5.11)$$

$$- \frac{1}{2} \int \langle (p_4 + p'_4)(G_4 + G'_4) A^k A'_k \rangle dx^{(3)} + \mathbf{v} \cdot \mathbf{p}^{(4)} + m_0 c^2 (1 - v^2/c^2)^{\frac{1}{2}},$$

where the canonical momentum $\mathbf{p}^{(4)}$ is given by

$$\mathbf{p}^{(4)} = \frac{\partial L}{\partial \mathbf{v}} = m_0 \mathbf{v} (1 - v^2/c^2)^{-\frac{1}{2}} + \frac{e_0}{c} \mathbf{A}(\mathbf{x}_{(t)}). \quad (5.12)$$

With the help of (5.8) the momentum \mathbf{P} defined in (5.5) may be written in the similar form

$$\mathbf{P} = \frac{1}{4c} \int \langle (\mathbf{p} + \mathbf{p}') (G_4 + G'_4) A^k A'_k \rangle dx^{(3)} + \frac{e_0}{c} \mathbf{A}(\mathbf{x}_{(t)}) + m_0 \mathbf{v} (1 - v^2/c^2)^{-\frac{1}{2}}. \quad (5.13)$$

Another expression for the energy-momentum four-vector is obtained from (5.2) and (5.5) by substitution from (5.4) and rearrangement of the terms in the following way:—

$$P_1 = P'_1 + m_0 v_1 (1 - v^2/c^2)^{-\frac{1}{2}} + K_1,$$

$$cP'_1 = \frac{1}{2} \hbar^{-2} \int \langle (\dot{p}_k p^k \delta_1^4 - p^4 p_1 - p^4 p'_1) B^k B'_k \rangle dx^{(3)} + e_0 \tilde{A}_1(\mathbf{x}_s),$$

$$cK_1 = \hbar^{-2} \int \left\langle \left(\frac{e^{(P' - P)/\hbar^2} - 1}{P' - P} \right) (p_4 + p'_4) p'_1 P B^k B'_k \right\rangle dx^{(3)} + e_0 \{A_1(\mathbf{x}_s) - \tilde{A}_1(\mathbf{x}_s)\}, \quad (5.14)$$

where B_k is again the pseudo-potential $e^{-P/\hbar^2} A_k$, and

$$\tilde{A}_1(\mathbf{x}_{(t)}) = \int B_1(x) e^{P/\hbar^2} \delta(\mathbf{x} - \mathbf{x}_{(t)}) dx^{(3)}. \quad (5.15)$$

From this the expression

$$cP_4 = cP_4^f + \mathbf{v} \cdot (\mathbf{P} - \mathbf{P}^f) + m_0 c^2 (1 - v^2/c^2)^{1/2} + (cK_4 - \mathbf{v} \cdot \mathbf{K}) \quad (5.16)$$

for the Hamiltonian energy follows at once. It will now be shown that

$$P_4^f = \frac{1}{c} \left[\int \mathbf{E} \wedge \mathbf{H} dx^{(3)}, \quad \frac{1}{2} \int (\mathbf{E}^2 + \mathbf{H}^2) dx^{(3)} \right],$$

$$\mathbf{E} = - \left(\frac{1}{c} \dot{\mathbf{B}} + \frac{\partial B_4}{\partial \mathbf{x}} \right), \quad \mathbf{H} = \frac{\partial}{\partial \mathbf{x}} \wedge \mathbf{B}. \quad (5.17)$$

Considering first the fourth component, one has from (5.14)

$$cP_4^f = \int \left\{ \frac{1}{2} \left(\frac{1}{c^2} \dot{\mathbf{B}}^2 - \mathbf{B} \cdot \Delta \mathbf{B} - \frac{1}{c^2} \dot{B}_4^2 + B_4 \Delta B_4 \right) + B_4 \square B_4 \right\} dx^{(3)}, \quad (5.18)$$

since $\square B_4 = e^{-P/\hbar c} \delta(\mathbf{x} - \mathbf{x}_{(t)})$. After the elimination of the time derivatives of B_4 with the help of the auxiliary condition $\frac{1}{c} \dot{B}_4 = -\frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{B}$, this reduces to

$$cP_4^f = \frac{1}{2} \int \left\{ \left(\frac{1}{c^2} \dot{\mathbf{B}}^2 - \frac{2}{c} B_4 \frac{\partial}{\partial \mathbf{x}} \cdot \dot{\mathbf{B}} - B_4 \Delta B_4 \right) - \mathbf{B} \cdot \Delta \mathbf{B} - \left(\frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{B} \right) \right\} dx^{(3)}, \quad (5.19)$$

which, by a trivial integration by parts, gives the formula (5.17). Similarly

$$cP^f = \int \left\{ -\frac{1}{c} \left(\frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{B} \right) \cdot \dot{\mathbf{B}} + \frac{1}{c} \frac{\partial B_4}{\partial \mathbf{x}} \dot{\mathbf{B}} + \mathbf{B} \square B_4 \right\} dx^{(3)}$$

$$- \int \left[\frac{1}{c} \left\{ \left(\frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{B} \right) \cdot \dot{\mathbf{B}} + \mathbf{B} \frac{\partial}{\partial \mathbf{x}} \cdot \dot{\mathbf{B}} \right\} + \left\{ \frac{\partial B_4}{\partial \mathbf{x}} \frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{B} + \mathbf{B} \Delta B_4 \right\} \right] dx^{(3)}, \quad (5.20)$$

which leads to the vector component of (5.17).

It may be shown further that the term $cK_4 - \mathbf{v} \cdot \mathbf{K}$ vanishes if terms in the acceleration and higher derivatives of \mathbf{v} are neglected. For then, with the help of (5.8),

$$\hbar^{-1} \int \left\langle \left(\frac{e^{(P' - P)/\hbar c} - 1}{P' - P} \right) (\rho_4 + \rho'_4) (\rho'_1 - \rho_1) P B^k B'_k \right\rangle dx^{(3)}$$

$$= \delta_1^4 \hbar^{-1} \int \left\langle (e^{(P' - P)/\hbar c} - 1) P B^k B'_k \right\rangle dx^{(3)} = -\delta_1^4 v^k \{A_k(\mathbf{x}_{(t)}) - A_k(\mathbf{x}_{(t)})\}, \quad (5.21)$$

so that

$$v^i K_i = \hbar^{-1} \int \left\langle (\rho_4 + \rho'_4) \left(\frac{e^{(P' - P)/\hbar c} - 1}{P' - P} \right) v^i \rho_1 P B^k B'_k \right\rangle dx^{(3)}, \quad (5.22)$$

which vanishes under the condition stated, since

$$v^i p_i P A_k = i \hbar e_0 e^{P/2\hbar} v^i \frac{\partial}{\partial x^i} \{v_k \delta(\mathbf{x} - \mathbf{x}_{(t)})\} = 0.$$

It is usual to express the Hamiltonian in terms of a set of canonical co-ordinates and their conjugate momenta. The co-ordinates most convenient for the present purpose are those of the electron, and the Fourier components $b_i(\mathbf{k})$ of $B_i(x)$, defined by

$$B_i(x) = c\Omega^{-1} \sum_{\mathbf{k}} b_i(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}/\hbar},$$

$$b_i(\mathbf{k}) = \frac{\Omega^{\frac{1}{2}}}{(2\pi\hbar)^{\frac{3}{2}}} \int B_i(x) e^{-i\mathbf{k} \cdot \mathbf{x}/\hbar} d\Omega, \quad (5.23)$$

for a finite region Ω of real space. Expressed in terms of these co-ordinates, however, the Lagrangian (5.1) and all derived quantities, including the Hamiltonian, contain all even derivatives of $b_i(\mathbf{k})$ with respect to time. It is clear on physical grounds, however, that *the state of the system is completely determined given only the values of the co-ordinates and their first derivatives at any time.** All higher derivatives have therefore to be eliminated. The elimination of the high derivatives of b_i is easily effected with the help of the field equations, which can be written in the form

$$\delta_i(\mathbf{k}) + \frac{c^2 k^2}{\hbar^2} b_i(\mathbf{k}) = \frac{e_0 \Omega^{\frac{1}{2}} e^{-k^2/4\omega^2}}{(2\pi\hbar)^{\frac{3}{2}}} e^{-e^2 D^2/4c^2} (v_i e^{i\mathbf{k} \cdot \mathbf{x}_{(t)}/\hbar}), \quad D \equiv \frac{d}{dt}, \quad (5.24)$$

and from which one obtains by repeated differentiation

$$D^{2n} b_i = \left(\frac{ick}{\hbar}\right)^{2n} b_i + \frac{D^{2n} - \left(\frac{ick}{\hbar}\right)^{2n}}{D^2 - \left(\frac{ick}{\hbar}\right)^2} \frac{e_0 \Omega^{\frac{1}{2}} e^{-k^2/4\omega^2}}{(2\pi\hbar)^{\frac{3}{2}}} e^{-e^2 D^2/4c^2} (v_i e^{i\mathbf{k} \cdot \mathbf{x}_{(t)}/\hbar}). \quad (5.25)$$

The equation of motion of the electron, which must be employed to eliminate the high derivatives of \mathbf{x} , is of the form

$$\ddot{\mathbf{x}} = f(b_i, \dot{b}_i, \mathbf{x}, \dot{\mathbf{x}}, \dot{\psi}, \ddot{\psi}, \dots), \quad (5.26)$$

where f is a known function of its arguments. The method which can be applied is to write

$$\dot{\psi}^{(n+1)} = f(b_i, \dot{b}_i, \mathbf{x}, \dot{\mathbf{x}}, \dot{\psi}^{(n)}, \ddot{\psi}^{(n)}, \dots),$$

$$\ddot{\psi}^{(n+1)} = \frac{d}{dt} \dot{\psi}^{(n)}, \text{ etc.,}$$

* A rigorous mathematical proof of this is postponed for future consideration.

and to proceed step by step, assuming that in the first approximation the motion of the electron is uniform, so that $\dot{\mathbf{v}}^{(1)} = \ddot{\mathbf{v}}^{(1)} = \dots = 0$. Assuming for the present that the assertion italicized above is correct, the solution obtained in this way will be unique.

The effect of a procedure of the kind just described on the Hamiltonian, as shown in Appendix V, is to reduce the number of canonical co-ordinates and momenta to one for each degree of freedom of the system. At each step in the elimination procedure approximations

$$H^{(n)}(\delta_i, \dot{\delta}_i, \mathbf{x}, \dot{\mathbf{x}}, \psi^{(n)}, \dot{\psi}^{(n)}, \dots), \quad \mathbf{p}^{(n)}(\delta_i, \dot{\delta}_i, \mathbf{x}, \dot{\mathbf{x}}, \psi^{(n)}, \dot{\psi}^{(n)}, \dots)$$

to the Hamiltonian and momentum of the electron are obtained which for large n converge to their exact values. Since $\mathbf{v}, \dot{\mathbf{v}}, \dots$ occur always in the combinations $a\dot{\mathbf{v}}/c^2, a^2\ddot{\mathbf{v}}/c^2, \dots$, however, $H^{(1)}$ and $\mathbf{p}^{(1)}$ are sufficiently exact for all ordinary purposes. To this approximation, as has already been seen, the term $cK_4 - \mathbf{v} \cdot \mathbf{K}$ disappears from the formula (5.16), so that

$$H = cP_4 = m_0c^2(1 - v^2/c^2)^{1/2} + \mathbf{v} \cdot (\mathbf{P} - \mathbf{P}') + P_4', \quad (5.27)$$

Also, according to (3.5), with $F(P) = e_0e^{-P/\hbar\omega}$,

$$e^{P/\hbar\omega}\delta(\mathbf{x} - \mathbf{x}_{(t)}) = \frac{1}{(2\pi\hbar)^3} \int \frac{e^{-i\mathbf{k} \cdot (\mathbf{x} - \mathbf{x}_{(t)})/\hbar} d\mathbf{k}}{e^{-\{(\mathbf{k} \cdot \mathbf{v}/c)^2 - k^2\}/\hbar\omega}} \quad (5.28)$$

In the next approximation it would be necessary to use (3.23); it is clear, however, that the improvement in accuracy so obtained would be very slight. Substituting, therefore, (5.28) with (5.23) into (5.15), one has

$$\tilde{A}_i(\mathbf{x}_{(t)}) = c\Omega^{-1} \sum_{\mathbf{k}} e^{\{(\mathbf{k} \cdot \mathbf{v}/c)^2 - k^2\}/\hbar\omega} b_i(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{x}_{(t)}/\hbar} \quad (5.29)$$

to a high degree of approximation.

6. THE QUANTUM ELECTRODYNAMICS

It is now possible to effect a natural transition from the classical theory of the preceding sections to the corresponding quantum formalism. The commutation rules for the electromagnetic field, established in § 4 of Q.T.R., are

$$\frac{1}{i}\hbar(G^4 + G^4')[A_k(x), A_{k'}(x')] = g_{kk'}\delta(\mathbf{x} - \mathbf{x}'), \quad (t' = t), \quad (6.1)$$

whilst those for the electron are simply

$$[x_{(t)}^k, p_l^{(0)}] = i\hbar\delta_l^k \quad (k, l = 1, 2, 3), \quad (6.2)$$

where $\mathbf{p}^{(t)}$ is the *canonical* momentum defined in (5.12). According to (5.11) and (5.13) one has

$$P_l = \frac{1}{4c} \int \langle (p_l + \dot{p}_l)(G^4 + G^4') A^k A'_k \rangle dx^{(3)} + p_l^{(0)}, \quad (6.3)$$

$$p_l^{(0)} = m_0 c (1 - v^2/c^2)^{1/2} + \mathbf{v} \cdot \mathbf{p}^{(t)}.$$

Since $p_l^{(0)}$ is assumed to commute with A_k , it follows, as in § 4 of Q.T.R., that

$$[P_l, A_k(x)] = i\hbar \frac{\partial A_k}{\partial x^l}(x); \quad (6.4)$$

also

$$[P_l, x_{(t)}^k] = i\hbar\delta_l^k \quad (l = 1, 2, 3); \quad [P_4, \mathbf{x}_{(t)}] = i\hbar\mathbf{v}/c. \quad (6.5)$$

It is easy to see that the components of P_l will commute among themselves; for, according to (6.4) and (6.5), one has $[P_4, \mathbf{P}] = \frac{i\hbar}{c} \frac{d}{dt} \mathbf{P}$, which vanishes, according to the conservation law proved with the help of the field equations in the previous section; and the components of the cartesian vector \mathbf{P} involve no operators which do not mutually commute. Since $PA_k = e^{P/2\hbar} v_k \delta(\mathbf{x} - \mathbf{x}_{(t)})$ commutes with A_k , (6.1) reduces on substitution from (5.4) to the form

$$-\frac{1}{2}\hbar^{-2} e^{-(P+P')/4\hbar} (p_4 + p'_4) [A_k(x), A_l(x')] = g_{kl} \delta(\mathbf{x} - \mathbf{x}'), \quad (t' = t),$$

which is consistent with the usual commutation rule

$$[B_k(x), B_l(x')] = i\hbar g_{kl} \delta(\mathbf{x} - \mathbf{x}'), \quad (t' = t). \quad (6.6)$$

The latter may therefore be regarded as the fundamental commutation rule, from which (6.1) is readily inferred.

The commutators of the components of P'_l , expressed by (5.17), are readily evaluated in terms of

$$\begin{aligned} \tilde{\mathbf{H}}(\mathbf{x}_{(t)}) &= \int \mathbf{H}(x) e^{P/4\hbar} \delta(\mathbf{x} - \mathbf{x}_{(t)}) dx^{(3)}, \\ \tilde{\mathbf{E}}(\mathbf{x}_{(t)}) &= \int \mathbf{E}(x) e^{P/4\hbar} \delta(\mathbf{x} - \mathbf{x}_{(t)}) dx^{(3)}, \end{aligned} \quad (6.7)$$

in the following way:—

$$\left. \begin{aligned} [P'_1, P'_2] &= \iint [E_1 H_2 - E_2 H_1, E'_1 H'_1 - E'_1 H'_2] dx^{(3)} dx^{(3)'} \\ &\quad - i\hbar \iint \left(\frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{E} H_2 - E_2 \frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{H} \right) dx^{(3)} - i\hbar c_0 \tilde{H}_2, \\ [P'_2, P'_1] &= i\hbar c_0 \tilde{H}_1, \quad [P'_3, P'_1] = i\hbar c_0 \tilde{H}_3, \\ [P'_1, P'_4] &= \frac{1}{2} \iint [E_1 H_3 - E_3 H_1, \mathbf{E}'^2 + \mathbf{H}'^2] dx^{(3)} dx^{(3)'} \\ &\quad - i\hbar \iint \left(\frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{E} E_1 - H_1 \frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{H} \right) dx^{(3)} = -i\hbar c_0 \tilde{E}_1, \\ [P'_2, P'_4] &= -i\hbar c_0 \tilde{E}_2, \quad [P'_3, P'_4] = -i\hbar c_0 \tilde{E}_3. \end{aligned} \right\} \quad (6.8)$$

One has further

$$\left. \begin{aligned} [\tilde{B}_1, P'_1] &= \int [\tilde{B}_1, E'_1 H'_1 - E'_1 H'_2] dx^{(3)'} = 0, \\ [\tilde{B}_1, P'_2] &= i\hbar \tilde{H}_2, \quad [\tilde{B}_1, P'_3] = -i\hbar \tilde{H}_3, \\ [\tilde{B}_1, P'_4] &= \frac{1}{2} \int [\tilde{B}_1, \mathbf{E}'^2 + \mathbf{H}'^2] dx^{(3)'} = -i\hbar \tilde{E}_1. \end{aligned} \right\} \quad (6.9)$$

Hence the field energy P'_4 commutes with the vector $\mathbf{P}' - e \cdot \tilde{\mathbf{B}}$, which may be regarded as the momentum vector of the field, and satisfies

$$[P'_1 - e_0 \tilde{B}_1, P'_2 - e_0 \tilde{B}_2] = i\hbar c_0 \tilde{H}_3, \text{ etc.} \quad (6.10)$$

The total field A_k may be separated into the unperturbed self-field of the electron $A_k^{(0)}$, and a radiation field A_k^r . Then if the total momentum and energy, given by (5.14) in conjunction with (5.17), are expressed in the form

$$\left. \begin{aligned} c\mathbf{P} &= c\mathbf{p}^0 + \int \mathbf{E}^r \wedge \mathbf{H}^r dx^{(3)} - e_0 \tilde{\mathbf{A}}^r, \\ cP_4 &= cp_4^0 + \frac{1}{2} \int (\mathbf{E}^r{}^2 + \mathbf{H}^r{}^2) dx^{(3)} - e_0 \tilde{A}_4^r. \end{aligned} \right\} \quad (6.11)$$

p^0 and p_4^0 may be interpreted as the entire momentum and energy associated with the electron. On account of (6.5), the normal commutation rules

$$[p_k^0, x_{(l)}^0] = i\hbar \delta_{kl} \quad (k, l = 1, 2, 3); \quad [p_4^0, \mathbf{x}_{(l)}] = i\hbar \mathbf{v}/c \quad (6.12)$$

are satisfied. Substituting (6.10) into the formula (5.27) for the Hamiltonian, the expression

$$\begin{aligned} cp_4^0 &= m_0 c^2 (1 - v^2/c^2)^{1/2} + \mathbf{v} \cdot \left(\mathbf{p}_0 - \frac{1}{c} \int (\mathbf{E} \wedge \mathbf{H} - \mathbf{E}^r \wedge \mathbf{H}^r) dx^{(3)} - \frac{e_0}{c} \tilde{\mathbf{A}}^r \right) \\ &\quad + \frac{1}{2} \int \{ (\mathbf{E}^2 + \mathbf{H}^2) - (\mathbf{E}^r{}^2 + \mathbf{H}^r{}^2) \} dx^{(3)} + e_0 \tilde{A}_4^r \end{aligned} \quad (6.13)$$

for p_4^0 results. It will now be shown that on writing $\mathbf{E} = \mathbf{E}^r + \mathbf{E}^{(i)}$, $\mathbf{H} = \mathbf{H}^r + \mathbf{H}^{(i)}$, in this expression,

$$\int \{(\mathbf{E}^{(i)} \cdot \mathbf{E}^{(r)} + \mathbf{H}^{(i)} \cdot \mathbf{H}^r) - \mathbf{v} \cdot (\mathbf{E}^{(i)} \wedge \mathbf{H}^r + \mathbf{E}^r \wedge \mathbf{H}^{(i)})/c\} dx^{(3)} = 0 \quad (6.14)$$

and

$$\int \{ \frac{1}{2} (\mathbf{E}^{(i)2} + \mathbf{H}^{(i)2}) - \mathbf{v} \cdot (\mathbf{E}^{(i)} \wedge \mathbf{H}^{(i)})/c \} dx^{(3)} = m_0 c (1 - v^2/c^2)^{\frac{1}{2}}, \quad (6.15)$$

so that an analogue

$$cp_4^0 = m(1 - v^2/c^2)^{\frac{1}{2}} + \mathbf{v} \cdot \left(\mathbf{p}_0 - \frac{e_0}{c} \tilde{\mathbf{A}}^r \right) + e_0 \tilde{A}_4^r, \quad m = 2m_0 \quad (6.16)$$

of Dirac's equation for the electron is obtained, with a mass m just *twice* the value m . The formula

$$m = \left(\frac{2}{\pi} \right)^{\frac{1}{2}} \frac{e_0^2}{4\pi a c^2} = \left(\frac{2}{\pi} \right)^{\frac{1}{2}} \frac{e^2}{a c^2} \quad (e_0 = (4\pi)^{\frac{1}{2}} e) \quad (6.17)$$

for the mass of the electron found in Q.T.R. is thereby confirmed.

To prove (6.14), one observes from (3.10) that $B_i^{(i)}$ satisfies $\mathbf{B}^{(i)} = \frac{\mathbf{v}}{c} B_4^{(i)}$ and $\dot{B}_4^{(i)} = \mathbf{v} \cdot \frac{\partial}{\partial \mathbf{x}} B_4^{(i)}$, so that

$$\left. \begin{aligned} \mathbf{E}^{(i)} &= -\frac{1}{c} \dot{\mathbf{B}}^{(i)} - \frac{\partial B_4^{(i)}}{\partial \mathbf{x}} = \left(\frac{\mathbf{v} \cdot \mathbf{v}}{c} \cdot \frac{\partial B_4^{(i)}}{\partial \mathbf{x}} - \frac{\partial B_4^{(i)}}{\partial \mathbf{x}} \right); & \mathbf{H}^{(i)} &= \frac{\partial}{\partial \mathbf{x}} \wedge \mathbf{B}^{(i)} = -\frac{\mathbf{v}}{c} \wedge \frac{\partial B_4^{(i)}}{\partial \mathbf{x}}; \\ \mathbf{E}^{(i)} + \mathbf{v} \wedge \mathbf{H}^{(i)}/c &= -(1 - v^2/c^2) \frac{\partial B_4^{(i)}}{\partial \mathbf{x}}; & \mathbf{H}^{(i)} - \mathbf{v} \wedge \mathbf{E}^{(i)}/c &= 0. \end{aligned} \right\} \quad (6.18)$$

Applying these results to (6.14), the left-hand side reduces to $-(1 - v^2/c^2) \int \frac{\partial B_4^{(i)}}{\partial \mathbf{x}} \cdot \mathbf{E}^r dx^{(3)}$, which vanishes on integration by parts, since

$\frac{\partial}{\partial \mathbf{x}} \cdot \mathbf{E}^r = 0$. To prove (6.15) it is necessary only to notice that, according

to the calculations of § 3, the energy $\frac{1}{2} \int (\mathbf{E}^{(i)2} + \mathbf{H}^{(i)2}) dx^{(3)}$ and momentum $\frac{1}{c} \int \mathbf{E}^{(i)} \wedge \mathbf{H}^{(i)} dx^{(3)}$ of the unperturbed self-field of the electron reduce to $m_0 c^2 (1 - v^2/c^2)^{-\frac{1}{2}}$ and $m_0 \mathbf{v} (1 - v^2/c^2)^{-\frac{1}{2}}$ respectively. The validity of (6.16) is therefore verified.

The Hamiltonian energy obtained from (6.11) by the substitution of (6.16) is

$$H = m c^2 (1 - v^2/c^2)^{\frac{1}{2}} + \mathbf{v} \cdot \left(\mathbf{p}_0 - \frac{e_0}{c} \tilde{\mathbf{A}}^r \right) + \frac{1}{2} \int (\mathbf{E}^{r2} + \mathbf{H}^{r2}) dx^{(3)}. \quad (6.19)$$

Here the interaction term $-e_0 \mathbf{v} \cdot \mathbf{A}^*/c$ may justifiably be regarded as small and treated quantum-mechanically as a perturbation. Before the transition to quantum theory is completed, however, some questions connected with the correspondence principle have to be considered. Firstly, Dirac's matrices α , β may be introduced by writing

$$(1 - v^2/c^2)^{1/2} \rightarrow \psi^* \beta \psi; \quad \mathbf{v} \rightarrow \psi^* \boldsymbol{\alpha} \psi \quad (6.20)$$

in the Hamiltonian. The velocity, however, occurs also in the interaction term in a more fundamental way, for, according to (5.29),

$$\tilde{\mathbf{A}}_r = c\Omega^{-1} \sum_{\mathbf{k}} (\mathbf{B}^*(\mathbf{k}) \exp(-i\mathbf{k} \cdot \mathbf{x}_e/\hbar) + \mathbf{B}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}_e/\hbar)) \cdot \exp\{(\mathbf{k} \cdot \mathbf{v}/c)^2 - k^2/4b^2\},$$

$$\mathbf{b}(\mathbf{k}) = \mathbf{B}^*(-\mathbf{k}) + \mathbf{B}(\mathbf{k}), \quad (6.21)$$

where $\mathbf{B}(\mathbf{k})$ may be supposed to contain the positive frequencies, and $\mathbf{B}^*(-\mathbf{k})$ the negative frequencies of the Fourier component $\mathbf{b}(\mathbf{k})$. Here it would be impossible to substitute from (6.20), and in applying the correspondence principle one therefore writes

$$\mathbf{v} \rightarrow \mathbf{p}_1^0/(m^2 c^2 + p^{02})^{1/2}. \quad (6.22)$$

This is strictly correct for the unperturbed electron, and it would be pointless to introduce a correction for the interaction here, as terms in $\alpha \mathbf{v}/c$ are already omitted from the Hamiltonian.

The order of the factors in (6.21) has been deliberately chosen to secure an expression which is hermitian and at the same time conforms with the usual rule that the factor $\mathbf{B}^*(\mathbf{k})e^{-i\mathbf{k} \cdot \mathbf{x}_e/\hbar}$ representing negative frequencies shall be written first, and the factor $\mathbf{B}(\mathbf{k})e^{i\mathbf{k} \cdot \mathbf{x}_e/\hbar}$, representing positive frequencies, last. The same rule is commonly applied to the term

$$\frac{1}{2} \int (\mathbf{E}_r^2 + \mathbf{H}_r^2) d\mathbf{x}^{(3)} = \sum_{\mathbf{k}} \left\{ \dot{\mathbf{B}}^*(\mathbf{k}) \cdot \dot{\mathbf{B}}(\mathbf{k}) + \frac{c^2 k^2}{\hbar^2} \mathbf{B}^*(\mathbf{k}) \cdot \mathbf{B}(\mathbf{k}) \right\}, \quad (6.23)$$

to prevent the appearance of the divergent "zero-point" energy $\sum_{\mathbf{k}} \frac{1}{2} c \hbar k$.

The Hamiltonian for an electron in interaction with a radiation field is then

$$H = \sum_{\mathbf{k}} \left(\dot{\mathbf{B}}^* \cdot \dot{\mathbf{B}} + \frac{c^2 k^2}{\hbar^2} \mathbf{B}^* \cdot \mathbf{B} \right) + c\psi^* \{ m c \beta + \boldsymbol{\alpha} \cdot (\mathbf{p}_0 - \frac{e_0}{c} \tilde{\mathbf{A}}_r) \}, \quad (6.24)$$

with $\tilde{\mathbf{A}}_r$ given by (6.21).

The problem of the interaction of a number of electrons with the radiation field may be treated similarly, and it is well known that such a problem may be reduced to the single-body problem through the introduction of second quantization of the electron field as well as of the

radiation field. The above results show that the customary theory is modified only by the presence of a factor $\exp\{(\mathbf{k} \cdot \mathbf{p}_0)^2 / (m^2 c^2 + p_0^2) - k^2\} / 4b^2$ in the matrix element representing the absorption or emission of a photon with momentum \mathbf{k} by the electron. Here $[p_0, (m^2 c^2 + p_0^2)^{1/2}]$ is to be interpreted as the energy-momentum of the electron in the *initial* state when the photon is *emitted*, in the *final* state when it is *absorbed*.

Thus the effect of the substitution of the reciprocally invariant Lagrangian operator for the usual one in the theory of the interaction of electrons and photons is summarized in the following simple prescription: to the matrix elements of the perturbation energy representing the emission of a photon with momentum \mathbf{k} by an electron with momentum p_0 , or the absorption of a photon with momentum \mathbf{k} by an electron with momentum $\mathbf{p}_0 - \mathbf{k}$, join a factor $\exp\{(\mathbf{k} \cdot \mathbf{p}_0)^2 / (m^2 c^2 + p_0^2) - k^2\} / 4b^2$. It will be found that this prescription makes many of the divergent terms of ordinary quantum electrodynamics finite.

APPENDIX I

PROOF OF THE RELATION

$$G(s) = \sum_0^s \frac{(-)^m 2^m m!}{(s-m)! (2m+1)!} = \frac{1}{(2s+1)s!} \quad (i)$$

Let

$$g_m(s) = \frac{(-)^m 2^m m!}{(s-m)! (2m+1)!},$$

then

$$G(s) = \sum_{m=0}^s g_m(s). \quad (ii)$$

Writing $g_m(s)$ in terms of $g_m(s+1)$, one has

$$\begin{aligned} G(s) &= \sum_0^s g_m(s+1) \left\{ \left(s + \frac{3}{2}\right) - \left(m + \frac{1}{2}\right) \right\} \\ &= \{G(s+1) - g_{s+1}(s+1)\} \left(s + \frac{3}{2}\right) + G(s) - g_s(s) - \frac{1}{2(s+1)!} \end{aligned}$$

or

$$G(s+1) \left(s + \frac{3}{2}\right) = g_s(s) + \left(s + \frac{3}{2}\right) g_{s+1}(s+1) + \frac{1}{2(s+1)!} \quad (iii)$$

Substituting $g_m(s)$ into (iii), one has

$$G(s+1) = \frac{1}{(2s+3)(s+1)!}.$$

Hence by mathematical induction one obtains the required result.

APPENDIX II

PROOF OF THE RELATION

$$\sum_{m=0}^{\mu} \frac{\Gamma(m+\frac{1}{2})}{\Gamma(s+m+1)} = \frac{2}{1-2s} \left\{ \frac{\Gamma(\mu+\frac{3}{2})}{\Gamma(\mu+s+1)} - \frac{\Gamma(\frac{1}{2})}{\Gamma(s)} \right\}. \quad (i)$$

For $\mu=0$, it is easily seen that

$$\frac{\Gamma(\frac{1}{2})}{\Gamma(s+1)} = \frac{2}{1-2s} \left\{ \frac{\Gamma(\frac{3}{2})}{\Gamma(s+1)} - \frac{s\Gamma(\frac{1}{2})}{\Gamma(s+1)} \right\},$$

so that (i) is an identity.

Now the difference

$$\begin{aligned} \sum_{m=0}^{\mu+1} \frac{\Gamma(m+\frac{1}{2})}{\Gamma(s+m+1)} - \sum_{m=0}^{\mu} \frac{\Gamma(m+\frac{1}{2})}{\Gamma(s+m+1)} &= \frac{2}{1-2s} \left\{ \frac{\Gamma(\mu+\frac{3}{2})}{\Gamma(s+\mu+2)} - \frac{\Gamma(\mu+\frac{3}{2})}{\Gamma(\mu+s+1)} \right\} \\ &= \frac{2}{1-2s} \frac{\Gamma(\mu+\frac{3}{2})}{\Gamma(\mu+s+2)} \{ \mu+\frac{3}{2} - (\mu+s+1) \} \\ &= \frac{\Gamma(\mu+\frac{3}{2})}{\Gamma(\mu+s+2)}, \end{aligned}$$

identically. The relation is therefore established.

APPENDIX III

The correction is obtained by the substitution of A_1^k given by (4.32), (4.33) and (4.35) into the following expression:—

$$-e_0(\mathbf{E}_1 + \mathbf{v} \wedge \mathbf{H}_1/c) = e_0 \left(\frac{1}{c} \frac{\partial}{\partial t} \mathbf{A}_1 + \frac{\partial}{\partial \mathbf{x}} \psi_1 - \mathbf{v} \wedge \frac{\partial}{\partial \mathbf{x}} \wedge \mathbf{A}_1/c \right). \quad (i)$$

The terms which are proportional to the zeroth power of r in U^k give for (i)

$$\frac{1}{2} e_0^2 a c^{-4} (2\pi)^{-3/2} \frac{d}{dt} \left\{ \frac{d^3}{ds^3} x^k + \frac{1}{4c^2} \left(\frac{d^1}{ds^2} x^1, \frac{d^2}{ds^2} x^1 \right) \frac{dx^k}{ds} \right\}, \quad k=1, 2, 3, \quad (ii)$$

while the terms which are proportional to r give

$$\begin{aligned} & \frac{1}{2} e_0^2 a c^{-4} (2\pi)^{-3/2} \left[\frac{1}{c^2} (\mathbf{v} \cdot \mathbf{v}) \ddot{\mathbf{v}} + \frac{1}{c^2} \mathbf{v} \wedge \mathbf{v} \wedge \ddot{\mathbf{v}} + \frac{1}{2} 3 c^{-2} \gamma^2 (\dot{\mathbf{v}} \cdot \mathbf{v}) \dot{\mathbf{v}} + \frac{1}{2} 5 c^{-2} \gamma^2 (\dot{\mathbf{v}} \cdot \mathbf{v}) (\dot{\mathbf{v}} \cdot \mathbf{v}) \mathbf{v} \right. \\ & \quad + \frac{1}{c^2} (\dot{\mathbf{v}} \cdot \mathbf{v}) \left\{ \gamma^2 \ddot{\mathbf{v}} + \frac{1}{2} 5 c^{-2} \gamma^2 (\mathbf{v} \cdot \mathbf{v}) \ddot{\mathbf{v}} + \frac{1}{2} 5 c^{-2} \gamma^2 (\dot{\mathbf{v}} \cdot \dot{\mathbf{v}}) \mathbf{v} + \frac{15}{8 c^2} (\dot{\mathbf{v}} \cdot \mathbf{v}) \dot{\mathbf{v}} \right. \\ & \quad + \frac{3 \cdot 5 \cdot 7}{8 c^4} \gamma^2 (\dot{\mathbf{v}} \cdot \mathbf{v})^2 \dot{\mathbf{v}} + \frac{5}{4 c^2} \gamma^2 (\ddot{\mathbf{v}} \cdot \mathbf{v}) \mathbf{v} + \frac{1}{2} 5 c^{-2} \gamma^2 \dot{\mathbf{v}} \cdot \mathbf{v} \mathbf{v} \\ & \quad \left. \left. + \frac{5 \cdot 7}{2 c^4} \gamma^2 (\dot{\mathbf{v}} \cdot \mathbf{v}) (\dot{\mathbf{v}} \cdot \mathbf{v}) \mathbf{v} + \frac{3 \cdot 5 \cdot 7}{8 c^4} \gamma^2 (\dot{\mathbf{v}} \cdot \dot{\mathbf{v}}) (\dot{\mathbf{v}} \cdot \mathbf{v}) \mathbf{v} + \frac{5 \cdot 7 \cdot 9}{8 c^4} \gamma^2 (\dot{\mathbf{v}} \cdot \mathbf{v})^2 \mathbf{v} \right\} \right] \\ & = \frac{1}{2} e_0^2 a c^{-4} (2\pi)^{-3/2} \frac{d}{dt} \left\{ -\frac{1}{4} \frac{d^3 \mathbf{x}}{ds^3} + \frac{7}{8 c^2} \frac{d\mathbf{x}}{ds} \left(\frac{d^2 \mathbf{x}^k}{ds^2} \cdot \frac{d^2 \mathbf{x}_k}{ds^2} \right) \right\}. \end{aligned} \quad (iii)$$

Summing (2) and (3), one has the total correction

$$\frac{1}{2} e_0^2 a c^{-4} (2\pi)^{-3/2} \frac{d}{dt} \left\{ \frac{3}{4} \frac{d^3 \mathbf{x}}{ds^3} + \frac{9}{8 c^2} \left(\frac{d^2 \mathbf{x}^k}{ds^2} \cdot \frac{d^2 \mathbf{x}_k}{ds^2} \right) \frac{d\mathbf{x}}{ds} \right\}. \quad (iv)$$

APPENDIX IV

In this appendix consideration is given to a type of Lagrangian $L(x, x_1, x_2, \dots)$ depending on all time derivatives $x_1 = \dot{x}$, $x_2 = \ddot{x}$, \dots of a co-ordinate x , but leading to a field equation

$$\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial x_1} + \frac{d^2}{dt^2} \frac{\partial L}{\partial x_2} - \dots, \quad (i)$$

which has the property that its solution is uniquely determined by the values of x and \dot{x} only at any instant. As is well known, the Hamiltonian

$$\left. \begin{aligned} H &= -L + p_0 x_1 + p_1 x_2 + p_2 x_3 + \dots, \\ p_n &= \frac{\partial L}{\partial x_{n+1}} - \frac{d}{dt} \frac{\partial L}{\partial x_{n+2}} + \frac{d^2}{dt^2} \frac{\partial L}{\partial x_{n+3}} - \dots \end{aligned} \right\} \quad (ii)$$

is defined in such a way as to satisfy the conservation law $dH/dt = 0$ by virtue of the field equation (i). The field equation can then be reproduced by the elimination of the momenta p_n from the Hamiltonian equations

$$\dot{x} = \frac{\partial H}{\partial p_0} = x_1; \quad \dot{p}_0 = -\frac{\partial H}{\partial x} = \frac{\partial L}{\partial x} \quad (iii)$$

and

$$\dot{x}_n = \frac{\partial H}{\partial p_n} = x_{n+1}; \quad \dot{p}_n = -\frac{\partial H}{\partial x_n} = \frac{\partial L}{\partial x_n} - \dot{p}_{n-1} \quad (n = 1, 2, \dots). \quad (\text{iv})$$

For this purpose the x_n have to be treated as independent variables, but as it is here supposed that only x and x_1 are independent, it must be possible to express the Hamiltonian in terms of x and p only, where p is the momentum conjugate to x . Assume, then, that $x_n = x_n(x, p)$, so

$$\left. \begin{aligned} \frac{\partial H}{\partial x} = & - \left(\frac{\partial L}{\partial x} + \frac{\partial L}{\partial x_1} \frac{\partial x_1}{\partial x} + \frac{\partial L}{\partial x_2} \frac{\partial x_2}{\partial x} + \dots \right) \\ & + \dot{p}_0 \frac{\partial x_1}{\partial x} + \dot{p}_1 \frac{\partial x_2}{\partial x} + \dots + \frac{\partial p_0}{\partial x} x_1 + \frac{\partial p_1}{\partial x} x_2 + \dots \end{aligned} \right\} \quad (\text{v})$$

Eliminating $\frac{\partial L}{\partial x}$, $\frac{\partial L}{\partial x_1}$, etc. with the help of (iii) and (iv), one has

$$\frac{\partial H}{\partial x} = - \left(\dot{p}_0 + \dot{p}_1 \frac{\partial x_1}{\partial x} + \dot{p}_2 \frac{\partial x_2}{\partial x} + \dots \right) + \frac{\partial p_0}{\partial x} x + \frac{\partial p_1}{\partial x} x_1 + \dots \quad (\text{vi})$$

On replacing \dot{p}_n by $\frac{\partial p_n}{\partial x} \dot{x} + \frac{\partial p_n}{\partial p} \dot{p}$ and \dot{x}_n by $\frac{\partial x_n}{\partial x} \dot{x} + \frac{\partial x_n}{\partial p} \dot{p}$, the terms in \dot{x} cancel, leaving

$$\frac{\partial H}{\partial x} = - \dot{p} \left\{ \left(\frac{\partial x}{\partial x} \frac{\partial p_0}{\partial p} - \frac{\partial x}{\partial p} \frac{\partial p_0}{\partial x} \right) + \left(\frac{\partial x_1}{\partial x} \frac{\partial p_1}{\partial p} - \frac{\partial x_1}{\partial p} \frac{\partial p_1}{\partial x} \right) + \dots \right\} \quad (\text{vii})$$

In a similar way it may be shown that

$$\frac{\partial H}{\partial p} = \dot{x} \left\{ \left(\frac{\partial x}{\partial x} \frac{\partial p_0}{\partial p} - \frac{\partial x}{\partial p} \frac{\partial p_0}{\partial x} \right) + \left(\frac{\partial x_1}{\partial x} \frac{\partial p_1}{\partial p} - \frac{\partial x_1}{\partial p} \frac{\partial p_1}{\partial x} \right) + \dots \right\} \quad (\text{viii})$$

Then, if

$$[x, p_0] + [x_1, p_1] + [x_2, p_2] + \dots = 1, \quad [a, b] = \frac{\partial a}{\partial x} \frac{\partial b}{\partial p} - \frac{\partial a}{\partial p} \frac{\partial b}{\partial x}, \quad (\text{ix})$$

the Hamiltonian equations

$$\frac{\partial H}{\partial x} = -\dot{p}, \quad \frac{\partial H}{\partial p} = \dot{x} \quad (\text{x})$$

are satisfied simultaneously.

The relation (ix) is therefore the generalization of the usual commutation rule for such systems; it is, in fact, equivalent to the relation $[x, p] = 1$ which is normally associated with the equations (x).

REFERENCES TO LITERATURE

- BOPP, F., 1940. *Ann. d. Phys.*, XXXVIII, 345.
—, 1943. *Ibid.*, XLII, 575.
BORN, M., 1909. *Ann. d. Phys.*, XXX, 1.
—, 1934. *Proc. Roy. Soc.*, A, CXLIII, 410.
—, 1939. *Proc. Roy. Soc. Edin.*, A, LIX, 219.
BORN, M., and GREEN, H. S., 1949. *Proc. Roy. Soc. Edin.*, A, LXII, 470.
BORN, M., and INFELD, L., 1934. *Proc. Roy. Soc.*, A, CXLIV, 425.
BORN, M., and RUMER, G., 1931. *Zeits. f. Phys.*, LXIX, 141.
DIRAC, P. A. M., 1938. *Proc. Roy. Soc.*, A, CLXVII, 148.
—, 1942. *Ibid.*, CLXXX, 1.
HEITLER, W., 1944. *The Quantum Theory of Radiation*, Oxford University Press.
HEITLER, W., and PENG, H. W., 1942. *Proc. Camb. Phil. Soc.*, XXXVIII, 296.
HERGLOTZ, G., 1903. *Nach. K. Ges. Wiss. (Göttingen)*, 357.
KRAZER, A., 1904. *Verh. des 3. Internationalen Math.-Kongresses (Heidelberg)*.
McMANUS, H., 1948. *Proc. Roy. Soc.*, A, CLXXXXV, 323.
PAIS, A., 1948. *Developments in the Theory of the Electron*, Publication of Princeton Institute for Advanced Studies.
PODOLSKY, B., 1942. *Phys. Rev.*, LXII, 68.
PRYCE, M. H. L., 1938. *Proc. Roy. Soc.*, A, CLXVIII, 389.
SCHWINGER, J., 1948. *Phys. Rev.*, LXXIII, 416.
—, 1949. *Ibid.*, LXXV, 651.
SOMMERFELD, A., 1904. *Nach. K. Ges. Wiss. (Göttingen)*.
WEISSKOPF, V., 1934. *Zeits. f. Phys.*, LXXXIX, 27.

(Issued separately May 14, 1951)

XI.—Application of Relaxation Methods to Compressible Flow past a Double Wedge.* By **A. R. Mitchell**, Ph.D., and **D. E. Rutherford**, Dr.Math., D.Sc., United College, University of St Andrews. (With Seven Text-figures.)

(MS. received July 18, 1949. Revised MS. received February 4, 1950.
Read February 27, 1950).

SYNOPSIS

The relaxation technique of R. V. Southwell is shown to be applicable in certain cases to transonic problems. For a uniform stream with a low subsonic velocity impinging on a symmetrical 2-dimensional double wedge, an asymmetrical supersonic region can be isolated in the neighbourhood of the corner of the wedge, and the streamlines and the values of the Mach number within this supersonic region can be determined with the aid of relaxation methods. Difficulties must be expected to occur in the neighbourhood of the sonic line, but in the present problem these have been surmounted.

I. INTRODUCTION

ALTHOUGH considerable progress has recently been made in the hodograph plane, it seems unlikely that analytical solutions will be obtained in the near future of problems concerning a compressible fluid, in which both subsonic and supersonic regions occur. Considerable importance must therefore still be attached to numerical methods and to their development. The results described in the present paper are part of a wider investigation undertaken by one of us (A. R. Mitchell) with a view to discovering how far the relaxation technique devised by R. V. Southwell can be extended to problems of the type described above. The present paper treats mixed flow past a double wedge by relaxation methods. All the numerical calculations were carried out by A. R. Mitchell

2. THE RELAXATION METHOD

The equations of 2-dimensional steady motion for the irrotational flow of a non-viscous compressible fluid may be written in the following form (Green and Southwell, 1943):—

$$\nabla^2(\chi\psi) - \psi\nabla^2\chi = 0, \quad (1)$$

$$\frac{2}{\gamma-1} \left\{ 1 - \left(\frac{\chi}{\chi_0} \right)^{\frac{2}{\gamma-1}} \right\} = \frac{\chi^2}{c_s^2} \left\{ \left(\frac{\partial\psi}{\partial x} \right)^2 + \left(\frac{\partial\psi}{\partial y} \right)^2 \right\}. \quad (2)$$

* Assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

In these formulæ, ψ denotes the stream function; χ stands for ρ^{-1} , ρ being the density; γ denotes the ratio of the specific heats; and c is the local speed of sound. c_s , χ_s are the values of c , χ respectively at a stagnation point.

It will be advantageous for our purposes to write these equations in the non-dimensional forms (Green and Southwell, 1943):

$$\nabla^2(\chi\psi) - \psi\nabla^2\chi = 0, \quad (3)$$

$$\frac{2}{\gamma-1}\{1 - \chi^{2(1-\gamma)}\} = \chi^2 \left\{ \left(\frac{\partial\psi}{\partial x} \right)^2 + \left(\frac{\partial\psi}{\partial y} \right)^2 \right\}. \quad (4)$$

To do so, we select some significant linear dimension h pertaining to the specific problem under consideration, and write

$$\nu = \frac{G}{\rho_s c_s h},$$

in which G is the mass flow per second under free stream conditions. Changing our notation slightly by writing χ , ψ , x , y for the non-dimensional quantities χ/χ_s , ψ/G , x/h , y/h respectively, equations (1) and (2) take the non-dimensional forms (3) and (4)

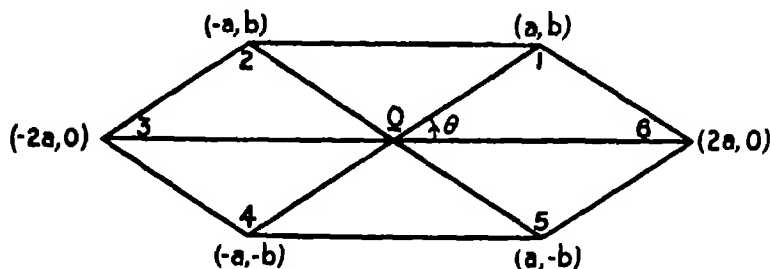


FIG. 1

In order to use the method of relaxation the whole field of the fluid is covered by a regular network. Although in this paper we shall only use a network of equilateral triangles, we think that it is worth while pointing out that such a network is only a particular case of an isosceles triangular net. A typical group of six isosceles triangles is illustrated in fig. 1. By varying the angle θ , the shape of such an isosceles triangle may be varied to suit the requirements of a particular problem. It will be seen that the square and triangular nets described by Southwell are particular cases ($\theta = 45^\circ$, 60°) of this more general net. The effectiveness of this network, however, may be impaired if θ has a value outside the range $45^\circ < \theta < 75^\circ$, the reason being that in such cases the six nodes of a typical hexagon, shown in fig. 1, are not the six nodes of the net which are closest to the centre of the hexagon.

The finite difference approximation to equation (3) applicable to a net of isosceles triangles may be shown to be

$$F_0 \equiv \frac{1}{\delta^2} \sum_{i=1,2,4,6} \chi_i (\psi_i - \psi_0) + \frac{1}{2} \left(\frac{1}{a^2} - \frac{1}{\delta^2} \right) \sum_{i=3,5} \chi_i (\psi_i - \psi_0) = 0, \quad (5)$$

in which χ_i, ψ_i denote the values of χ, ψ at the node labelled i ($i=0, 1, 2, \dots, 6$) in fig. 1. It can also be shown that the approximate equation corresponding to (4) is

$$\begin{aligned} \frac{\gamma-1}{\gamma-1} (1 - \chi_0^{2(1-\gamma)}) - \frac{\nu^2 \chi_0^4}{\delta^2} \left\{ \psi_0^2 + \frac{1}{4} \sum_{i=1,2,4,6} \psi_i (\psi_i - 2\psi_0) \right\} \\ + \frac{1}{4} \left(\frac{1}{a^2} - \frac{1}{\delta^2} \right) \nu^2 \chi_0^4 \left\{ \psi_0^2 + \frac{1}{2} \sum_{i=3,5} \psi_i (\psi_i - 2\psi_0) \right\}. \end{aligned} \quad (6)$$

To obtain an approximate solution for any particular problem, the finite difference equations (5) and (6) must be satisfied approximately throughout the given region, either by trial and error, or by a mathematical plan called by Southwell (1946) a *pattern*. In practice, the least laborious method is to employ a combination of both methods.

Initially, a value for ψ is assumed at each node of the net. Substituting the appropriate values into equation (6), a relation is obtained from which χ may be determined at any node. Equation (5) now determines the *residual* F at each node. The procedure thereafter is to modify repeatedly the ψ -distribution, either by trial and error or according to a recognized plan, so that the residuals at each node are made as small as possible.

In the case of the equilateral triangular net, which is the appropriate one for the problem here investigated, $b = a\sqrt{3}$. Writing $A = 2a = 2b/\sqrt{3}$, and defining R by the relation

$$R = \chi^{-4} \{1 - \chi^{2(1-\gamma)}\},$$

equations (5) and (6) become respectively

$$F_0 = \sum_{i=1,2,4,5,6} \chi_i (\psi_i - \psi_0) = 0, \quad (7)$$

$$R_0 = \frac{(\gamma-1)\nu^2}{A^2} \left\{ \psi_0^2 + \frac{1}{6} \sum_{i=1,2,3,4,5,6} \psi_i (\psi_i - 2\psi_0) \right\}. \quad (8)$$

We shall find it convenient to express the relationship between R and χ in the form $\chi = f(R)$, the graph of which is exhibited in fig. 2. It is easily verified that

$$\frac{d\chi}{dR} = f'(R) = \frac{\chi^5}{2\{(1+\gamma)\chi^{2(1-\gamma)} - 2\}}. \quad (9)$$

This relationship is graphed in fig. 3a and fig. 3b for $\gamma = 1.4$.

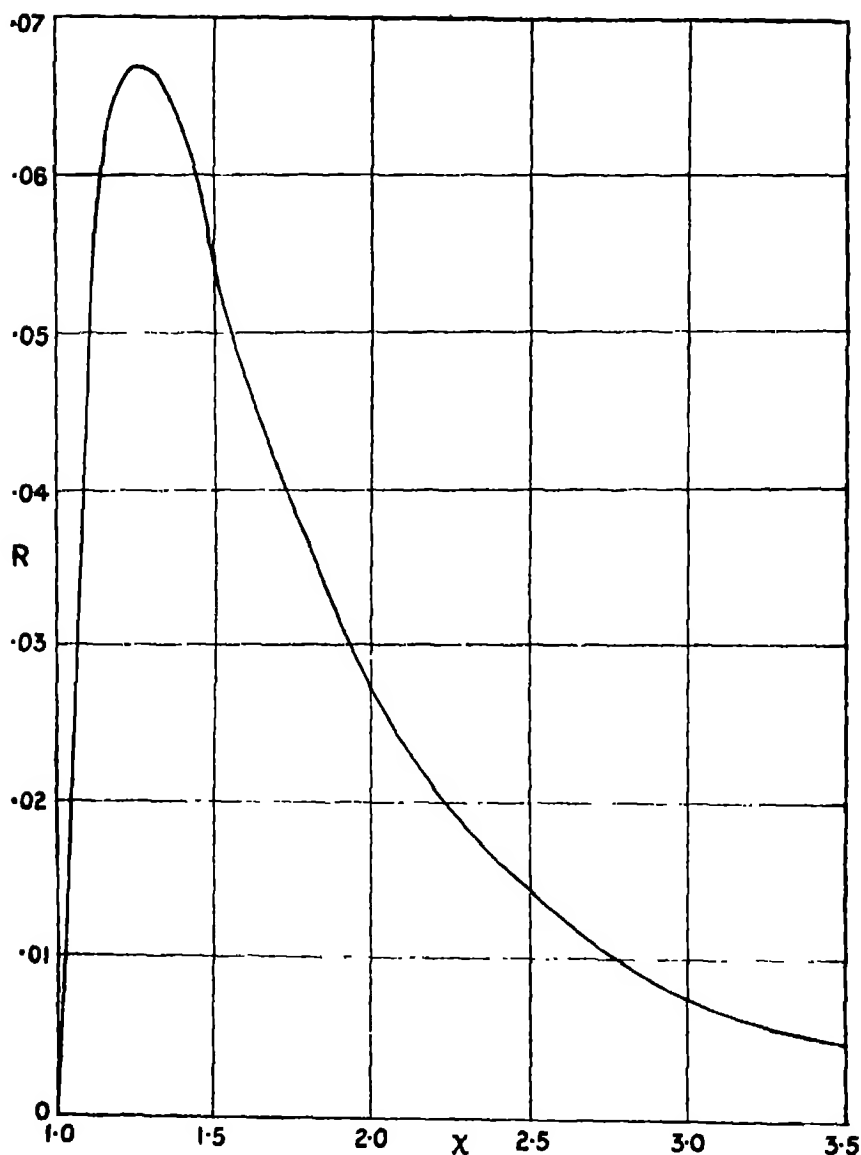


FIG. 2

Our next task in working out the pattern is to determine the changes $\delta F_0, \delta F_1, \dots, \delta F_6$ consequent on an assumed modification $\delta\psi_0$ of ψ_0 . The required formulæ, whose derivation will not be given here, are

$$\delta F_0 = - \left[\sum_{i=1}^6 \frac{(\gamma_i - 1) \nu_i^2}{3A^2} \{f'(R)\}_i (\psi_0 - \psi_i)^2 + \sum_{i=1}^6 \chi_i \right] \delta\psi_0, \quad (10)$$

$$\delta F_j = \left[\chi_0 + \frac{(\gamma-1)\nu^2}{A^2} \{f'(R)\}_0 (\psi_0 - \psi_i) \left(2\psi_0 - \frac{1}{3} \sum_{i=1}^6 \psi_i \right) \right] \delta \psi_0, \quad (j=1, \dots, 6). \quad (11)$$

These formulæ are obtained in the same manner as the corresponding formulæ for the square network were obtained by Green and Southwell (1943). As we have said, the original values ψ being known at each node,

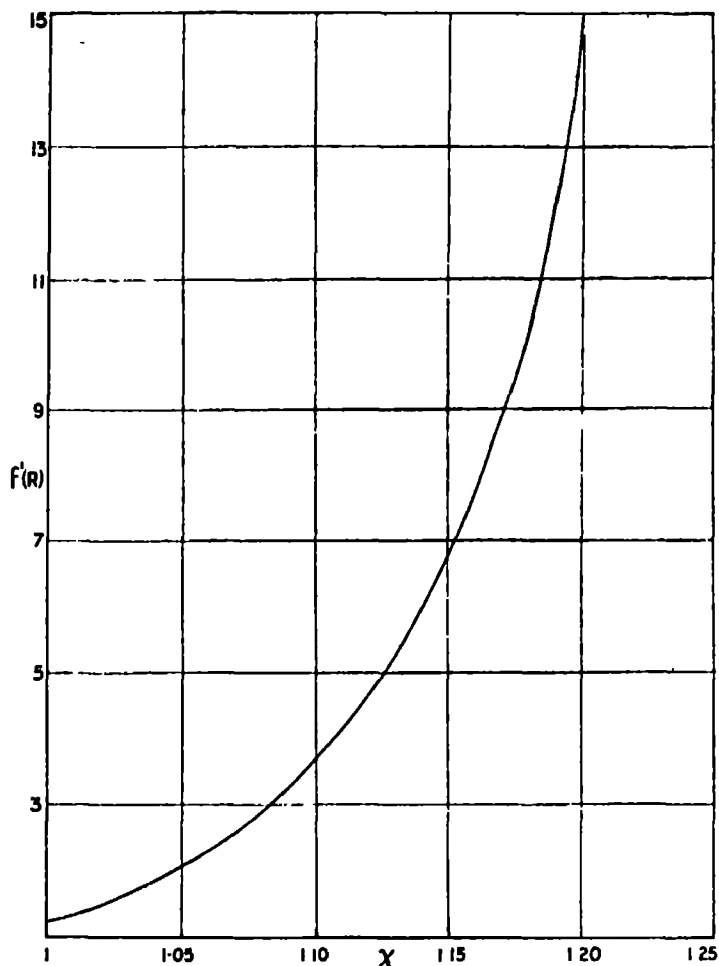


FIG. 3a.—Subsonic range

χ can be determined. Accordingly $f'(R)$ can be determined from figs. 3a and 3b at each node. From the formulæ (10) and (11), the modifications $\delta F_0, \dots, \delta F_6$ arising out of an increment $\delta \psi_0$ in ψ_0 may then be evaluated. By a judicious choice of $\delta \psi_0$, the residual F_0 may be almost eliminated, although this will in general be accompanied by a slight increase in the residuals at the surrounding nodes.

When this process has been effected at all nodes at which outstanding residuals occur, the resulting ψ -distribution is taken as a basis for a net of finer mesh. The whole procedure is then repeated until the residuals remaining may be considered negligible.

In the problem described in § 5 there would seem to be an inherent difficulty in the employment of the relaxation pattern in the neighbourhood of the wedge apex, since χ has a singularity at the apex. Since ψ is known at all the nodes on the surface of the wedge, there is no difficulty

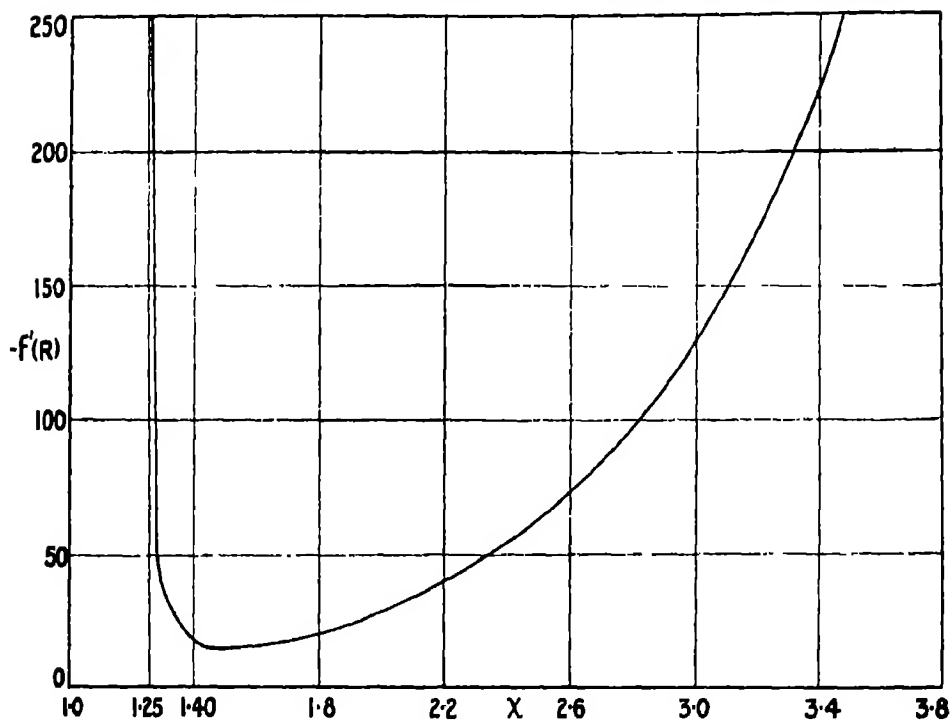


FIG. 36.—Supersonic range

in determining χ from equation (8) at all interior nodes, including the layer of nodes nearest to the boundary. We cannot, however, determine F_0 at these "nearest" nodes, since formula (7), when applied to such nodes, would require a knowledge of χ on the boundary itself. On the other hand, F_0 can be obtained from (7) at the "next nearest" nodes to the boundary. The relaxation pattern can accordingly be employed to determine an improved value for ψ at all interior nodes with the exception of those nearest to the boundary.

Since the "next nearest" layer can be brought as close as we please to the boundary by choosing a fine enough mesh, the singularity of χ at the apex need cause no concern. It should be borne in mind, however, that

the resulting approximate values of ψ at the "nearest" nodes will be less accurate than those at other nodes, for the degree of approximation obtained at "nearest nodes" will be that corresponding to a mesh half as fine as the mesh actually used.

3. SUPERSONIC FLOW

It has been suggested by other authors (Fox and Southwell, 1944; Southwell, 1946; Emmons, 1944) that certain problems in supersonic flow may not be amenable to relaxation treatment. In this section we shall attempt to show that many problems in supersonic flow can be examined just as easily as in the subsonic case by the relaxation technique. To explain the nature of Southwell's reluctance to use the method of relaxation when the flow is supersonic, it is necessary to refer to the formulæ (9), (10) and (11).

If we tabulate the relation (9) for the case $\gamma = 1.4$ as under,

χ	1.00	1.15	1.20	1.25 -	1.25 +	1.50	2.00	2.70	3.50
M	0	0.75	0.88	1.0 -	1.0 +	1.4	1.9	2.4	3.0
$f'(R)$	1.25	6.70	15.5	+ ∞	- ∞	- 14	- 26	- 77	- 236

it will be observed that for supersonic speeds ($M > 1$), $f'(R)$ has large negative values, large in comparison with the positive values which it has for subsonic speeds. From the occurrence of $f'(R)$ in formulæ (10) and (11), Southwell concluded that for a Mach number M greater than unity very large modifications δF_j ($j=0, 1, \dots, 6$) would be occasioned by small increments $\delta\psi_0$, and that consequently the approximations yielded by the relaxation method would not be convergent, or would, at best, converge very slowly. The discontinuity in $f'(R)$ at $M=1$ clearly indicates that the relaxation method will break down at sonic velocity, but in several problems which we have attempted no insuperable difficulty was found, except in the immediate neighbourhood of the sonic line, in applying the relaxation technique to supersonic problems with $M < 3$.

It will be observed that in formula (10) there are two terms which contribute to the value of δF_0 . So long as the term involving $f'(R)$ is not of a greater order of magnitude than the term $\sum_{i=1}^6 \chi_i$, the order of magnitude of $f'(R)$ itself need not cause concern. In such problems as we have

examined, it has been found that the term in $f'(R)$ did not outweigh the other term.

For example, at one node in whose surrounding hexagon a Mach number of 0.9 occurred, we found that

$$\sum_i \frac{(\gamma-1)v^2}{3A^2} \{f'(R)\}_i (\psi_0 - \psi_i)^2 = +1.19,$$

$$\sum_i \chi_i = +6.82,$$

and consequently that

$$\delta F_0 = -8.01.$$

At another node in whose surrounding hexagon a Mach number of 1.05 occurred, we obtained

$$\sum_i \frac{(\gamma-1)v^2}{3A^2} \{f'(R)\}_i (\psi_0 - \psi_i)^2 = -15.82,$$

$$\sum_i \chi_i = +8.23,$$

$$\delta F_0 = +7.59.$$

Analogous results were obtained for each δF_i . It may be added that the two examples given above were specifically chosen as the most unfavourable cases in the problem under examination for the finest net used. It should be pointed out, however, that the employment of a still finer net would have produced less favourable nodes, in whose surrounding hexagons $0.9 < M < 1.05$.

In conclusion, it would seem that although in certain supersonic problems the relaxation technique may become quite unworkable on account of very large values of $f'(R)$, yet in many cases this is not so. There are certainly many problems in supersonic flow which are amenable to relaxation methods.

4. METHODS AVAILABLE NEAR A SONIC LINE

It will be evident from what has been said in the last section that, on account of the singularity of $f'(R)$ at $\chi=1.25$ ($M=1$), the relaxation technique, as described above, must break down in the neighbourhood of the sonic line. As the sonic line is approached from either side, $f'(R)$ takes increasingly large positive or negative values. Equations (10) and (11) show that under such circumstances negligible changes in the assumed ψ -distribution give large changes in the residues. In consequence, the relaxation pattern becomes unworkable as a computational problem

The difficulty mentioned above may be seen in another way. For low subsonic or high supersonic flow, the modification in ψ is made solely with a view to the elimination of the residuals F . In such cases the appropriate value of χ can be read off from fig. 2, either a value less than 1.25 in the subsonic case or one greater than this value in the supersonic case. As, however, the sonic line is approached from either side, it will be found in general that a modification of ψ may result in a value of R which is greater than the maximum possible one of 0.067. In such a case (*cf.* fig. 2) no real value can be attached to χ . The problem therefore resolves itself into finding a modification of the ψ -distribution, which, while producing values of R which at each node are less than 0.067, at the same time eliminates the residuals F .

The sonic line is initially, in point of fact, an undetermined boundary between the subsonic and supersonic regions. Its position can only be determined by approaching it from either side asymptotically, by employing smaller and smaller hexagons, and by modifying the ψ -distribution subject to the stipulations concerning R already mentioned. This was eventually accomplished by a method of trial and error. So far we have been unable to devise any pattern which may be used in the immediate neighbourhood of the sonic line. The usual relaxation pattern will certainly break down in such regions.

Whether or not the procedure described above is a feasible one in every problem we are unable to say, but in several transonic problems which have been examined no insuperable difficulty has presented itself.

5. PROBLEM OF THE WEDGE

The particular problem described in this paper is that of a 2-dimensional compressible, irrotational, non-viscous flow in a channel, past a symmetrical double wedge, whose profile is shown in fig. 4. For simplicity in calculation the semi-angle of the wedge was taken to be 30° .

In the free stream a low Mach number $M=0.205$ ($\nu=0.200$) was chosen with a view to investigating how a supersonic region would develop around the apex P , even when the free stream speed was low subsonic. In addition, the choice of such a low free stream speed emphasizes the fact that fine details of the flow, in particular the flow around P , may be examined by reducing the mesh of the net appropriately.

The channel width $2OR$ was chosen to be five times the wedge width $2OP$. With this ratio the solution of the problem should be a close approximation to that of the flow past a wedge in a free stream. There is

A further refinement $A = 1/480$ was made incorporating the methods suggested in § 4, and by this means a provisional supersonic region around P was isolated. It was found, however, that large residuals persisted on the line of symmetry PR , and that these residuals could not be eliminated by any of the methods which we have described.

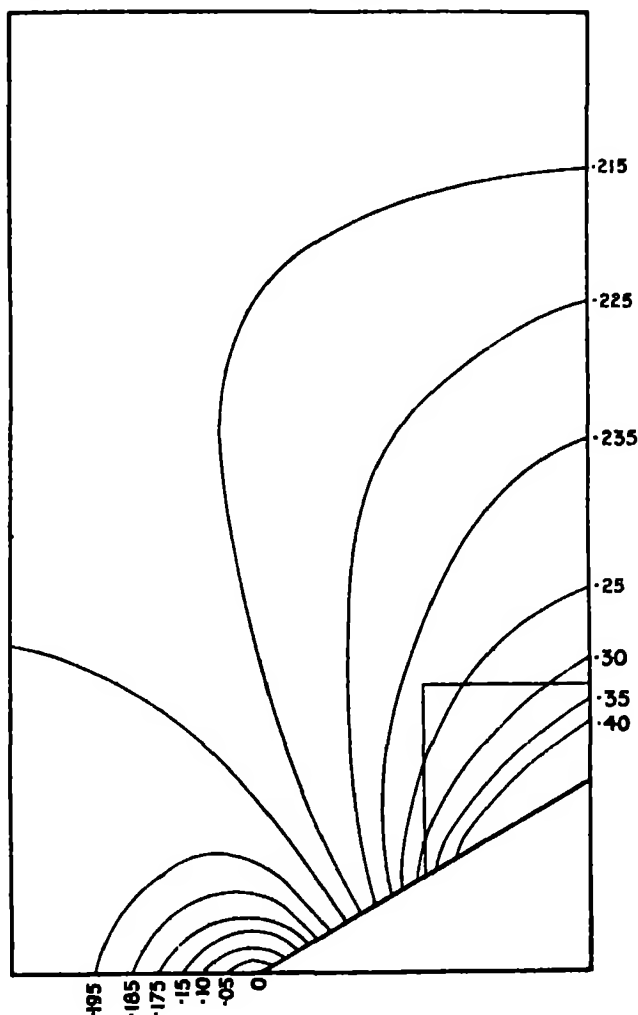


FIG. 5.—Lines of constant Mach number

At this stage, therefore, the assumption of symmetry was abandoned, but we retained, as a basis for further calculations, the portion upstream of the sonic line of the subsonic ψ -distribution which had already been calculated. The flow picture up to the line of symmetry is shown in figs. 5 and 6

After much manipulation, a ψ -distribution was obtained stretching downstream from the line of symmetry PR . This was done in a manner similar to that employed upstream of the line of symmetry. The residuals,

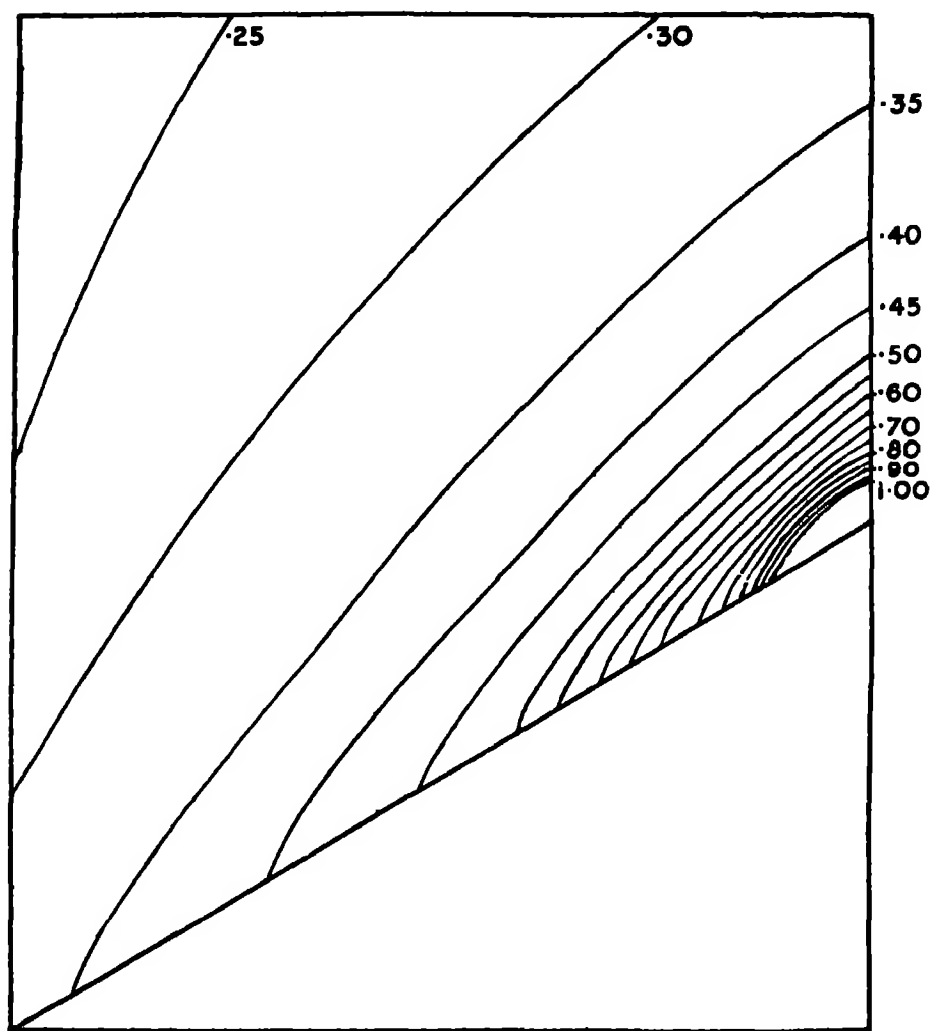


FIG. 6.—Lines of constant Mach number

including those on the line of geometrical symmetry PR , were now everywhere small. Fig. 7 shows the asymmetrical supersonic region in the vicinity of the wedge apex

It might also be pointed out that the relaxation methods here described may be used to determine the flow characteristics in a supersonic region, should they be required. At every node in the supersonic region the local

Mach number and flow direction are known. Consequently, the directions of the two characteristics may be calculated at each node, since the angle between either characteristic and the direction of flow at a point in a supersonic region is given by $\sin^{-1}(1/M)$. By interpolation, the directions of the two characteristics at every point of the supersonic region may be found. Accordingly, by starting anywhere on the boundary of the supersonic region, the two characteristics at the point, one of each family, can be traced throughout their length. This was in fact done, but in this problem no interesting features emerged.

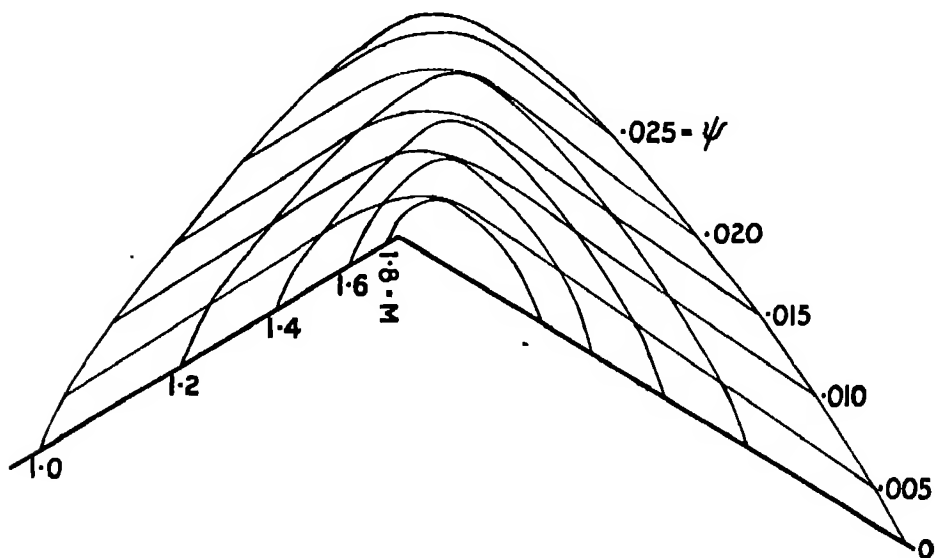


FIG. 7.—Supersonic region round wedge apex

It was not considered profitable to obtain a detailed solution for the subsonic region downstream of the line of symmetry, since its determination entailed no novel features.

It should be emphasized that we do not claim any great degree of accuracy for the portion of fig. 7 which lies between the stream-line $\psi = .010$ and the boundary of the wedge, nor for the very small corresponding region in fig. 6. The reason for this is explained at the end of § 2. Further, since the curvature of the stream-lines in this region is not obtainable with sufficient precision from our calculations, we do not know at what point the curvature of the lines $M = \text{constant}$ becomes pronounced so that they meet the boundary normally. That the lines $M = \text{constant}$ do behave in this way can be deduced from the formula $\tan \beta = R\dot{q}/q^2$, in which q is the speed, R is the radius of curvature of the stream-line and β is the angle between the stream-line and the line $M = \text{constant}$. The very

rapid increase in R as the boundary is approached accounts for the fact that β may change from, say, 40° to 90° in a very short distance.

6. SYMMETRY

In this section we shall discuss in greater detail the question of symmetry raised in § 5.

For the sake of argument, let us assume that the flow is symmetrical about the line PR . Since the velocity must be continuous at points on this line of symmetry, at such points the velocity must be parallel to the free stream and to the channel walls. Consider now a supersonic region surrounding the apex bounded by a line of constant Mach number M' where $M' > 1$. So long as the speed is increasing, the stream tubes must diverge. It is readily seen that it is impossible to insert stream-lines within this region, which diverge from this line $M = M'$ right up to the line of symmetry, and which at the same time cross the line of symmetry normally. Consequently, the assumption of a symmetrical flow implies that on any stream-line passing through the supersonic region there must be two points, symmetrically situated on either side of the line of symmetry, at which the speed attains a maximum value.

Although there are indications that this situation may arise in certain cases of flow about a blunt corner, it was considered unlikely that in our problem a maximum speed would occur in front of the line of symmetry.

The assumption of symmetry was therefore abandoned, and an asymmetrical solution was sought in which the maximum speed occurred downstream of the line PR .

There are additional arguments to support this point of view. Consider two points Z and Z' symmetrically situated within the supersonic region, with respect to PR . The situation at Z' downstream of Z is of course directly affected by the conditions existing at Z . On the other hand, the situation at Z can only be affected by that at Z' if the signal travels first of all downstream into the subsonic region, then travels backward within the subsonic region until it reaches a point upstream of Z , and finally travels downstream until it reaches Z . Thus, although the circumstances at each point can affect those at the other, they do so in essentially different ways. This argument indicates that it is at least probable that the flow pattern is asymmetrical.

Finally, in the well-known Meyer expansion, the flow is not symmetrical about the line of symmetry of the boundary. Since the flow in the immediate vicinity of the apex must resemble very closely a Meyer expansion, we conclude that the flow is certainly not symmetrical at the apex itself.

There are therefore good reasons for supposing that in mixed or supersonic flow a symmetrical solution will not be obtained, even when the boundary possesses a line of symmetry.

It was in fact found to be impossible to eliminate the residuals along PR when symmetry was assumed, and it was this fact which first led us to examine the assumption of symmetry more closely, and finally to discard it. It should be emphasized, however, that all the above objections to a symmetrical solution break down if the fluid is everywhere subsonic. As the free stream Mach number in our problem is very small, the presence of a supersonic region and its lack of symmetry may be attributed to the sharp corner at the apex of the boundary.

7. UNIQUENESS OF THE SOLUTION

So far as we are aware there has been no rigorous demonstration that there is a unique solution of the fundamental equations (1) and (2) which satisfies given boundary conditions.

Once the assumption of symmetry is abandoned in the present problem, the boundary conditions consist of $\psi=0$ along the channel axis and wedge sides, $\psi=1$ along the channel wall, and free stream conditions at entry to the channel, infinitely far from the wedge. Conditions of flow are not specified at any position downstream of the wedge. We are concerned with a mixed continuous flow, the only supersonic region being situated round the wedge apex.

Courant and Friedrichs (1948, p. 370) suggest that in such a mixed flow problem the above boundary conditions are sufficient, if continuity is assumed, to determine the flow uniquely, although there are some indications that a slight change in the shape of the wedge would alter the flow considerably. Certainly we were unable to obtain a solution other than, but differing slightly from, the solution obtained in § 5.

Thus, while we would hesitate to assert categorically that the solution exhibited in figs. 5, 6 and 7 is unique, we hold nevertheless that there are good grounds for believing this to be so.

8. CONCLUSION

The most significant feature of the foregoing study would seem to be the possibility of using the relaxation technique in certain problems of mixed and supersonic flow. Admittedly, difficulties will be encountered in the neighbourhood of the sonic line, and these must be overcome by the method outlined in § 4.

The possibility of treating supersonic problems carries with it the ability to determine graphically the flow characteristics by the methods of relaxation.

We should also draw attention once more to the asymmetrical character of the solution obtained. Like many other authors before us, we first of all jumped to the conclusion that a symmetrical solution could be found. Later we were forced to reject this hypothesis in favour of that of an asymmetrical solution.

Lastly, the isosceles triangular network described in § 2, while not utilized in the present paper, may be used to minimize computation in problems, in which the boundaries do not lie neatly on an equilateral triangular net. The isosceles network would, for example, have been advantageous in the present problem if the semi-wedge angle had been, say, 50° .

REFERENCES TO LITERATURE

- COURANT, R., and FRIEDRICHS, K. O., 1948. *Supersonic Flow and Shock Waves*, New York.
- EMMONS, H. W., 1944. "Numerical Solutions of Compressible Fluid Flow Problems", *A.R.C. Report* 7871.
- FOX, L., and SOUTHWELL, R. V., 1944. "On the Flow of Gas through a Nozzle with Velocities exceeding the Speed of Sound", *Proc. Roy. Soc., A*, C1XXXIII, 38-54.
- GREEN, J. R., and SOUTHWELL, R. V., 1943. "High Speed Flow of Compressible Fluid through a Two-Dimensional Nozzle", *Phil. Trans., A*, CCXXXIX, 367-386.
- JONES, M., BRIGHT, P., and ANDREWS, J. B., 1946. "The Squares Method applied to Compressible Flow with a Shock Wave", *R.A.E. Technical Note Aero* 1833.
- SOUTHWELL, R. V., 1946. *Relaxation Methods in Theoretical Physics*, Oxford.
- THOM, A., and KLANFER, L., 1947. "Some Arithmetical Studies of the Compressible Flow past a Body in a Channel", *A.R.C. Report* 11010.

XII.—Clebsch–Aronhold Symbols and the Theory of Symmetric Functions.* By **H. W. Turnbull** and **A. H. Wallace**,
The University, St Andrews.

(MS. received September 13, 1949. Revised MS. received January 25, 1950.
Read February 6, 1950)

SYNOPSIS

A square matrix $A = (a_{ij})$ is expressed symbolically in terms of Clebsch–Aronhold equivalent symbols $a_{ij} = a_i a_j = \beta_i \beta_j = \dots$, and the symbolic expressions for symmetric functions of the latent roots of A are considered, the relation between these functions and projective invariants of the bilinear form uAx being noted. The Newton and Brioschi relations between the symmetric functions are obtained by reduction of symbolic determinants and permanents respectively, and the Wronskian relations are shown to be equivalent to certain identities between determinants and permanents due to Muir. Also the fundamental theorem of symmetric functions is obtained symbolically as a consequence of the first fundamental theorem of invariants. The paper concludes with a note on the symbolization of the A -bialternants, that is of the traces of irreducible invariant matrices of A .

1. INTRODUCTION

THE Clebsch–Aronhold symbols were introduced by the authors from whom they derive their name in their papers on invariant theory (namely, Aronhold, 1858; Clebsch, 1861 a, b). Symbolic devices of a similar nature had already been used by Sylvester: in his paper “On the Principles of the Calculus of Forms” (Sylvester, 1852), symbols, which he called *umbræ*, were used in the discussion of commutants; while in another paper (Sylvester, 1851) a symbolic notation for the elements of a determinant is introduced which is identical with the notation used in the present paper to represent the elements of a matrix.

The usefulness of these symbols in various connections is due to the fact that a number of results in algebra are really substitutional in character, sets of suffices or variables being permuted in the various terms of the identities involved. But very often these sets of suffixes or variables contain repeats which made the manipulation of substitutional operators on them inconvenient. Clebsch–Aronhold symbols act as position-markers, and the substitutional operations may be transferred to them from the suffix—or variable—sets, while, by having a set of equivalent

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

symbolizations, it is ensured that the objects now being permuted contain no repeats.

As used in invariant theory, the symbols replace the coefficients of the ground forms by sets of symbolic vectors, and so the problem of finding all concomitants of a set of ground forms is reduced to that of finding all invariant polynomials in the elements of a set of vectors, either variables or symbolic coefficients, all of which undergo either a certain linear transformation or the transformation contragredient to it. And the first fundamental theorem of projective invariants shows that all such invariant functions are expressible rationally and integrally in terms of inner products of the type

$$u_e = \sum_{i=1}^m u_i x_i,$$

and bracket factors of the types

$$(u^{(1)} u^{(2)} \dots u^{(m)}) \quad \text{and} \quad (x^{(1)} x^{(2)} \dots x^{(m)}),$$

where u and the $u^{(i)}$ are cogredient row vectors, x and the $x^{(i)}$ are column vectors contragredient to them, and the bracket factors represent m -rowed determinants (Turnbull, 1945, pp. 173-180, and Chaps. XI and XII).

In particular the discussion of the invariant theory of a single form bilinear in two sets of m variables is given by Turnbull (1932), where the coefficients of the form

$$uAx = \sum_{i,j=1}^m u_i a_{ij} x_j$$

are written as

$$a_{ij} = \alpha_i \alpha_j = \beta_i \beta_j = \gamma_i \gamma_j = \dots$$

these all being equivalent symbolizations. In the course of the work certain results of the classical theory of symmetric functions are derived, and expressed, by means of the Clebsch-Aronhold symbols. It is the aim of this paper to push that investigation somewhat further.

In § 7 below use is made of a result in the theory of double standard forms which requires some explanation.

A Young tableau of shape (λ) consists of a set of n symbols, not necessarily all distinct, arranged in rows and columns, with λ_1 symbols in the first row, λ_2 in the second, λ_3 in the third, and so on, where

$$(\lambda) \equiv (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_k)$$

is a partition of n , such that

$$\begin{aligned} \lambda_1 + \lambda_2 + \dots + \lambda_k &= n, \\ \lambda_1 &> \lambda_2 > \dots > \lambda_k. \end{aligned}$$

Also the first column is fully occupied with its h symbols. For example,

corresponding to the partition $(3, 2, 2)$ is the shape $\begin{array}{ccc} & * & * & * \\ & * & * & \\ & * & * & \end{array}$. The

tableau is called standard if no column contains any symbol twice, and if, reading along each row from left to right or down each column, the symbols appear in some preassigned order.

If a set of r row vectors u, v, \dots, w and a set of the same number of column vectors x, y, \dots, z are given, then the r th compound inner product, or bideterminant, of the two sets of vectors, is defined as the r -rowed determinant

$$(uv \dots w | xy \dots z) \equiv \begin{vmatrix} u_x & u_y & \dots & u_z \\ v_x & v_y & \dots & v_z \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ w_x & w_y & \dots & w_z \end{vmatrix}.$$

If $r > m$, this vanishes.

If U is a tableau of shape (λ) whose symbols are row vectors, and X is a tableau of the same shape whose symbols are column vectors, the double form, written $\{U | X\}$, is defined to be the product of the compound inner products of the sets of vectors in the corresponding columns of U and X . Thus

$$\left\{ \begin{array}{ccc|ccc} u & u & v & x & y & z \\ v & w & & y & z & \\ w & & & z & & \end{array} \right\} = (uvw | xyz) (uw | yz) v_z.$$

If U and X are both standard, the form $\{U | X\}$ is called a double standard form.

It is proved in Turnbull (1945, Chap. XXIII) that double standard forms in a given set of variables are linearly independent, and any non-standard double form may be expressed linearly in terms of them. In this cited chapter, the notation for double forms does not involve curly brackets; they are introduced here to avoid confusion between the symbols of forms written side by side.

2. THE SYMBOLIZATION OF A SQUARE MATRIX

Let $A = [a_{ij}]$ be a square matrix of order $m \times m$ in which the i, j th element is symbolized by

$$a_{ij} = \alpha_i \alpha_j = \beta_i \beta_j = \gamma_i \gamma_j = \dots$$

The symbolic expression of a single element needs just one pair of symbols,

say α and a ; a product of any two elements needs two such pairs, say α , a and β , b , and so on; for example

$$a_{ij}a_{hk} = \alpha_i a_j \beta_h b_k = \beta_h b_k \alpha_i a_j;$$

but $\alpha_i a_j \alpha_h a_k$ is ambiguous. Each pair of symbols, as illustrated here, is equivalent to, and so interchangeable with, each other pair, English letter with English letter and Greek with Greek, such an interchange leaving unaltered the value of the expression which is symbolized. The symbolic factors are commutative with respect to multiplication, and of course this commutative property introduces no danger of ambiguous interpretation; in fact $a_{ij}a_{hk}$ could also be written as $\alpha_i \beta_h a_j b_k$, and in interpreting this symbolic product the factor α_i must be taken along with the a_j , even though these symbols do not actually appear together; the preassigned convention that α and a are to be regarded as companion symbols ensures an unambiguous reading of the symbolic product. We shall be dealing here with homogeneous polynomials in the elements a_{ij} of the matrix A , and the symbolization of each term of such a polynomial, say of degree r , will therefore contain r different English letters $a, b, c \dots$, and the corresponding r Greek letters α, β, γ .

It is worth noting that although, in general, $\alpha_i a_j$ and $a_i \alpha_j$ differ, being symbolizations of a_{ij} and a_{ji} respectively, they are equal if A is a symmetric matrix. Thus for a symmetric matrix the simpler notation

$$a_{ij} = a_i a_j = b_i b_j = c_i c_j = \dots$$

is available, without the use of Greek letters.

3. INVARIANTS UNDER THE TRANSFORMATION $\bar{A} = HAH^{-1}$

A homogeneous polynomial in the a_{ij} is a symbolic multilinear form in the elements of the column vectors

$$\alpha = \{\alpha_1 \alpha_2 \dots \alpha_m\}, \quad \beta = \{\beta_1 \beta_2 \dots \beta_m\}, \quad \dots$$

and the row vectors

$$a = [a_1 a_2 \dots a_m], \quad b = [b_1 b_2 \dots b_m], \quad \dots$$

Under the similarity transformation

$$\bar{A} = HAH^{-1},$$

which involves a non-singular matrix H of order $m \times m$, these symbolic vectors undergo the transformations

$$\left. \begin{aligned} \bar{\alpha} &= H\alpha, & \bar{\beta} &= H\beta, & \dots \\ \bar{a} &= aH, & \bar{b} &= bH, & \dots \end{aligned} \right\}, \quad (1)$$

where $\bar{a}, \bar{\beta}, \dots$ and \bar{a}, \bar{b}, \dots are formed, like the corresponding vectors a, β, \dots and a, b, \dots , from the symbolic form of \bar{A} , namely from

$$\bar{a}_{ij} = \bar{a}_i \bar{a}_j = \bar{\beta}_i \bar{b}_j = \dots$$

By the first fundamental theorem of projective invariants, any multilinear form $\Phi(a, \beta, \dots, a, b, \dots)$ in the elements of the symbolic vectors $a, \beta, \dots, a, b, \dots$, which is a nilbaric, or absolute, invariant under the transformations (1) for all non-singular H , that is a form such that

$$\Phi(\bar{a}, \bar{\beta}, \dots, \bar{a}, \bar{b}, \dots) = \Phi(a, \beta, \dots, a, b, \dots),$$

is always expressible as a polynomial in the inner products $a_a, a_b, \dots, b_a, b_b, \dots, c_a, c_b, \dots$, etc., where

$$a_a = \sum_{i=1}^m a_i a_i.$$

But if the elements of A are taken to be independent variables, it may be assumed that the latent roots of A are all distinct; and so any polynomial in the a_{ij} which is invariant under the transformation $\bar{A} = HAH^{-1}$ is first of all a polynomial in the latent roots of A , as may be seen by choosing H such that \bar{A} is a diagonal matrix. Then since the set of all non-singular matrices H includes the set of all permutation matrices, that is, matrices obtained from the m -rowed unit matrix by permuting its rows, any such invariant polynomial in the a_{ij} must be a symmetric function in the latent roots of A . Conversely, any symmetric function of the latent roots of A must be invariant under all transformations $\bar{A} = HAH^{-1}$, since such transformations leave unaltered the set of latent roots as a whole, at most permuting them among themselves.

Hence any symmetric function in the latent roots of A may be written symbolically as a polynomial in the inner products

$$a_a, a_b, \dots, b_a, b_b, \dots,$$

and conversely, any polynomial in the a_{ij} which, when symbolized, can be expressed rationally and integrally in terms of these inner products is a symmetric function of the latent roots of A .

4. CLOSED, OPEN, AND PRIME PRODUCTS

A symbolic product of the $a_a, a_b, \dots, b_a, b_b, \dots$, in which to each English letter appearing there is a corresponding Greek letter, and *vice versa*, will be called a *closed* product. A product such as $a_b b_y$ or $a_a b_y$, or a_b itself, where certain letters appear without their companions, will be called *open*. A closed product then symbolizes an actual polynomial in

the a_{ii} ; it is indeed a symmetric function of the latent roots of A ; but an open product is purely symbolic and has no actual interpretation in terms of the a_{ij} . A closed product which contains open factors only will be called *prime*; while if it can be factorized into further closed products it will be called *composite*. Thus each of $a_a, a_\beta b_a, a_\beta b_\gamma c_a, \dots$ is prime, whereas $a_a b_\gamma c_\beta$ is composite, having the closed factors a_a and $b_\gamma c_\beta$. The value of a prime product depends only on the number of distinct symbol pairs occurring—this is a direct consequence of the equivalence of symbol pairs; e.g. $a_\beta b_a = c_\beta d_\gamma$ and $a_\beta b_\gamma c_a = a_\gamma b_a c_\beta$. Hence the value of a composite product depends only on the way in which the total number of its distinct pairs of symbols is partitioned so as to form prime products.

This may be expressed differently by regarding a closed product involving the n letters a, b, c, \dots and their n companions $\alpha, \beta, \gamma, \dots$ as the result of operating on the sequence a, b, c, \dots in the product

$$a_a b_\beta c_\gamma \dots$$

with a certain permutation of the symmetric group on n letters; then the value of the composite product depends only on the class of the permutation in question. The product then contains a prime symbolic factor involving k symbol pairs, corresponding to each cycle of order k in the permutation. Let a prime product of k symbol pairs be denoted by s_k , and let

$$(\rho) = (\rho_1, \rho_2, \dots)$$

be a set of positive integers such that

$$\rho_1 + 2\rho_2 + 3\rho_3 + \dots = n.$$

Further, let

$$S_{(\rho)} = s_1^{\rho_1} s_2^{\rho_2} s_3^{\rho_3} \dots \quad (2)$$

It follows that

$$S_{(\rho)} = P_{(\rho)} a_a b_\beta c_\gamma \dots,$$

where $P_{(\rho)}$ is a permutation, of class (ρ) , operating on the sequence a, b, c, \dots . Or if $C_{(\rho)}$ is the sum of the permutations of class (ρ) in the symmetric group on n letters, and $k_{(\rho)}$ is the number of such permutations within the class, we may write

$$S_{(\rho)} = \frac{1}{k_{(\rho)}} C_{(\rho)} a_a b_\beta c_\gamma \dots \quad (3)$$

(The change from Frobenius' symbol $h_{(\rho)}$ to $k_{(\rho)}$ is made to avoid confusion, as $h_{(\rho)}$ is now used in a different connection.)

Thus any homogeneous symmetric function of degree n of the latent roots of A may be written symbolically as a linear combination of closed

products by the first fundamental theorem of invariants. Each of these products must be one of the $S_{(\rho)}$, and so may be written in the form (3). Thus the given symmetric function is the result of operating upon the sequence $a, b, c \dots$ in $a_\alpha b_\beta c_\gamma \dots$ with a linear combination of the operators $C_{(\rho)}$, that is, the function may be written as

$$X a_\alpha b_\beta c_\gamma \dots$$

where X is a substitutional operator and, moreover, is a linear combination of the class sums of the symmetric group on n letters. Equally well the operator X could act on the sequence $\alpha, \beta, \gamma, \dots$ of Greek letters instead of on the corresponding English ones.

5. CERTAIN SPECIAL SYMMETRIC FUNCTIONS

The prime products

$$s_1 = a_\alpha, \quad s_2 = a_\beta b_\alpha, \quad s_3 = a_\beta b_\gamma c_\alpha, \quad \dots$$

have already been introduced, but it remains to evaluate them in terms of the latent roots of A . To do this, consider the i, j th elements of the matrices $A, A^2, A^3 \dots$; they are, respectively,

$$a_i a_j; \quad a_i a_\beta b_j; \quad a_i a_\beta b_\gamma c_j; \quad \dots$$

by the ordinary rule for matrix multiplication. Putting $j=i$ and summing with respect to i , it follows that the traces of these matrices are, respectively,

$$a_\alpha; \quad a_\beta b_\alpha; \quad a_\beta b_\gamma c_\alpha; \quad \dots$$

that is,

$$s_r = \text{trace of } A^r = \text{tr } (A^r), \quad \text{say.}$$

By Sylvester's theorem, the trace of A^r is the sum of the r th powers of the latent roots of A , and so this gives the value of s_r .

Or, independently of Sylvester's theorem, s_r may be evaluated by using the fact that it is invariant under similarity transformations; thus A may be taken in the first place to be a diagonal matrix, with diagonal elements $\omega_1, \omega_2, \dots, \omega_m$. Then

$$a_{ij} = \delta_{ij} \omega_i.$$

In this case, s_3 (taking $r=3$ for convenience in writing) is given by

$$s_3 = a_\beta b_\gamma c_\alpha = \sum_{i,j,k} a_i \beta_i b_j \gamma_j c_k \alpha_k,$$

the summation being with respect to i, j and k independently, each from

1 to m . But

$$\sum_{i,j,k} a_i \beta_j b_k \gamma_i \epsilon_k a_k = \sum_{i,j,k} \delta_{ij} \omega_i \delta_{jk} \omega_j \delta_{ki} \omega_k \\ = \sum_i \omega_i^3.$$

And in general, as before,

$$s_r = \sum_{i=1}^m \omega_i^r.$$

Now consider the symbolic bideterminants

$$M_1 \equiv (a | a) \equiv a_a; \quad M_2 \equiv (ab | a\beta) \equiv \begin{vmatrix} a_a & a_\beta \\ b_a & b_\beta \end{vmatrix}; \\ M_3 \equiv (abc | a\beta\gamma) \equiv \begin{vmatrix} a_a & a_\beta & a_\gamma \\ b_a & b_\beta & b_\gamma \\ c_a & c_\beta & c_\gamma \end{vmatrix}; \text{ etc.,}$$

where M_r is an r th compound inner product or bideterminant. These are symmetric functions of the latent roots of A , for they are obtained from the product $a_a b_\beta c_\gamma \dots$ by determinantal permutation $\dot{a}_a \dot{b}_\beta \dot{c}_\gamma \dots$, and are hence sums of closed products. These M -invariants may be evaluated as in Turnbull (1932, p. 6), or again by supposing A to be diagonal, with $a_{ii} = \delta_{ii} \omega_i$. Then, for $r = 3$, say,

$$M_3 = \dot{a}_a \dot{b}_\beta \dot{c}_\gamma = \sum_{i,j,k} \dot{a}_i \dot{b}_j \dot{c}_k a_i \beta_j \gamma_k.$$

On account of the determinantal permutation (a sum of six terms) indicated by the dots over a , b and c , non-zero expressions are obtained in the summation with respect to i , j and k only if these indices are all distinct. Each set of three distinct indices appears in all possible orders, namely $3!$ times, and each time the only non-zero contribution is $\omega_i \omega_j \omega_k$. Hence

$$M_3 = 3! e_3,$$

where e_3 is the third elementary symmetric function of the latent roots of A . In general,

$$M_r = r! e_r,$$

where e_r is the r th elementary symmetric function of $\omega_1, \omega_2, \dots, \omega_m$.

A third set of symmetric functions is produced by permanental summation applied to $a_a b_\beta$, $a_a b_\beta c_\gamma$, \dots etc., namely

$$N_1 \equiv a_a, \\ N_2 \equiv \begin{vmatrix} + & + \\ a_a b_\beta & a_\beta b_a \end{vmatrix} \equiv a_a b_\beta + a_\beta b_a, \\ N_3 \equiv \begin{vmatrix} + & + & + \\ a_a b_\beta c_\gamma & a_\beta c_a \gamma & a_\gamma c_a \beta \end{vmatrix} \equiv a_a b_\beta c_\gamma + a_a c_\beta \gamma + a_\beta c_a \gamma + a_\beta b_\gamma c_a + a_\gamma c_a \beta + a_\gamma b_\beta c_a,$$

and so on. N_r will consist of $r!$ terms, each with a positive sign (whereas M_r consists of the same $r!$ terms, half with positive and half with negative signs). N_r may be called a bipermanent, and an appropriate notation is

$$N_2 = (ab \parallel a\beta),$$

$$N_3 = (abc \parallel a\beta\gamma),$$

and so on, where the double vertical line replaces the single line in M_r .

To evaluate the N_r , suppose, as in the case of the M_r , that A is diagonal. Then for $r=3$, say,

$$N_3 = \sum_{i,j,k} a_i \beta_j \gamma_k \mid a_i b_j c_k \mid$$

summing over all sets i, j, k , in all possible orders. In general, fixing attention on a certain set of r suffixes attached to the r letters a, β, γ, \dots , suppose that this set contains p equal suffixes of one kind, q equal suffixes of a second kind, s of a third kind, and so on. Then, in the summation with respect to sets of suffixes there are $\binom{r}{p, q, s, \dots}$ arrangements of this particular set. For each such arrangement the permanental summation with respect to a, b, c, \dots gives a non-zero contribution only when the arrangement of the suffixes of a, b, c, \dots coincides with that of the suffixes of a, β, γ, \dots (since A is diagonal); but since the sets of repeated suffixes may be permuted among themselves this coincidence may happen in $p!q!s! \dots$ ways. Hence the non-zero contribution to N_r due to this selected suffix set is $r!$ times a certain product of degree r of the latent roots, ω_i , of A , and the summation with respect to suffix sets becomes a summation over all possible products of degree r of the ω_i . In fact,

$$N_r = r! h_r,$$

where h_r is the r th complete homogeneous function of the latent roots of A .

6. RELATIONS BETWEEN SYMMETRIC FUNCTIONS

It has already been shown (Turnbull, 1932) that the M -invariants and the first m of the s_r each form a complete and irreducible set of invariants for a single bilinear form uAx , where u and x are contragredient vectors. Here the variable vector x is contragredient also to the symbolic vector a , and cogredient with the vector a . In the present paper these results, and certain others, will now be derived, the emphasis here, however, being put on the fact that the invariants discussed are symmetric functions of the

latent roots of a matrix, rather than on the fact that they are projective invariants of a certain bilinear form.

The theorems which are about to be proved on irreducible bases for symmetric functions give an illustration of the working of the second fundamental theorem of invariants, which states that every identity among invariants, expressed symbolically, can ultimately be expressed by means of certain fundamental identities together with the principle of interchange of equivalent symbols and the laws of ordinary algebra for the manipulation of symbolic inner products and bracket factors (Turnbull, 1945, p. 214). In the present instance these fundamental identities are equivalent to stating that any compound inner product M_r of order $r > m$ vanishes identically.

If, now, any compound inner product M_r is expanded as a determinant, there will be $(r-1)!$ terms, each with the same sign and each equal to s_r , since, in the symmetric group on r letters, there are just $(r-1)!$ permutations each consisting of one cycle of order r . The rest of the expansion is a polynomial in the s_i for $i < r$. If $r > m$, $M_r = 0$, and so, for $r > m$, the identity

$$0 = (r-1)! s_r + \phi(s_1, s_2, s_3, \dots, s_{r-1})$$

holds, ϕ being a polynomial. Hence s_r may be expressed as a polynomial in the functions s_i for $i < r$; and by repeating the process, an expression is obtained for s_r as a polynomial in the s_i for $i < m$. By the very nature of the symbolic expression of symmetric functions, any symmetric function of the latent roots of A may be written as a polynomial in the functions s_i , and by the result just obtained this polynomial may be reduced to a polynomial in the s_i for $i < m$. Alternatively any homogeneous symmetric function of degree n of the latent roots of A may be written as a linear combination of the functions

$$S_{(\rho)} = s_1^{\rho_1} s_2^{\rho_2} s_3^{\rho_3} \dots,$$

where ρ_r is zero whenever $r > m$.

In Turnbull (1932) it is shown that a symbolic reduction of the determinantal expression for M_r ($r < m$) leads to the formula

$$\frac{1}{(r-1)!} M_r = \frac{1}{(r-1)!} M_{r-1} s_1 - \frac{1}{(r-2)!} M_{r-2} s_2 + \dots + (-)^{r-1} s_r \dots \quad (4)$$

i.e.

$$r e_r = e_{r-1} s_1 - e_{r-2} s_2 + \dots + (-)^{r-1} s_r \dots \quad (5)$$

To prove this result, consider first the expansion of the symbolic determinant of r rows and r columns

$$(acde \dots | \beta\gamma\delta\epsilon \dots) = \begin{vmatrix} a_\beta & a_\gamma & a_\delta & a_\epsilon & \dots \\ c_\beta & c_\gamma & c_\delta & c_\epsilon & \dots \\ d_\beta & d_\gamma & d_\delta & d_\epsilon & \dots \\ e_\beta & e_\gamma & e_\delta & e_\epsilon & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}$$

in terms of its first column and cofactors; here each English letter except a has a companion Greek letter, and the English letter appears corresponding to each Greek letter except β . Now

$$\begin{aligned} (acde \dots | \beta\gamma\delta\epsilon \dots) &= a_\beta(cde \dots | \gamma\delta\epsilon \dots) - c_\beta(ade \dots | \gamma\delta\epsilon \dots) \\ &\quad + d_\beta(ace \dots | \gamma\delta\epsilon \dots) - e_\beta(acd \dots | \gamma\delta\epsilon \dots) + \dots \\ &= a_\beta(cde \dots | \gamma\delta\epsilon \dots) - c_\beta(ade \dots | \gamma\delta\epsilon \dots) \\ &\quad + c_\beta(ade \dots | \delta\gamma\epsilon \dots) - c_\beta(acd \dots | \epsilon\delta\gamma \dots) + \dots \end{aligned}$$

by the interchange of pairs of equivalent symbols. Then, by rearranging the rows and columns of the determinants on the right,

$$(acde \dots | \alpha\gamma\delta\epsilon \dots) = a_\beta(cde \dots | \gamma\delta\epsilon \dots) - (r-1)c_\beta(ade \dots | \gamma\delta\epsilon \dots) \quad (6)$$

where the cofactor of a_β is a sum of closed products, while the cofactor of c_β is similar to the original determinant, namely, with all the English and Greek letters paired off except the first in each set. Similarly if a is replaced by b ,

$$(bcde \dots | \beta\gamma\delta\epsilon \dots) = b_\beta(cde \dots | \gamma\delta\epsilon \dots) - (r-1)c_\beta(bde \dots | \gamma\delta\epsilon \dots),$$

i.e.

$$M_r = s_1 M_{r-1} - (r-1)c_\beta(bde \dots | \gamma\delta\epsilon \dots).$$

Then, by applying the reduction formula (6) to $(bde \dots | \gamma\delta\epsilon \dots)$,

$$\begin{aligned} M_r &= s_1 M_{r-1} - (r-1)c_\beta[b_\gamma(de \dots | \delta\epsilon \dots) - (r-2)d_\gamma(be \dots | \delta\epsilon \dots)] \\ &= s_1 M_{r-1} - (r-1)s_2 M_{r-2} + (r-1)(r-2)c_\beta d_\gamma(be \dots | \delta\epsilon \dots). \end{aligned}$$

And so, by repeated application of the reduction formula (6), and final division by $(r-1)!$, the equation (4), or the equivalent equation (5), is obtained.

The equations (5) for $r=1, 2, \dots, m$ may be solved to give s_1, s_2, \dots, s_m as polynomials in e_1, e_2, \dots, e_m . Hence any homogeneous symmetric function of degree n in the latent roots of A is expressible as a polynomial in the elementary symmetric functions e_1, e_2, \dots, e_m or, in other words, as

a linear combination of the double standard forms

$$e_{(\lambda)} = \{S_{(\alpha)}^{(\lambda)} \mid S_{(\beta)}^{(\lambda)}\},$$

where $S_{(\alpha)}^{(\lambda)}$ and $S_{(\beta)}^{(\lambda)}$ are tableaux of the same shape (λ) (with not more than m rows) formed from English and Greek letters, respectively, in which the letters appear in alphabetical order, reading along each row in turn from left to right. As an illustration of the last statement, the product $e_1 e_2^2 e_3$ may be expressed symbolically as

$$\frac{1}{3! 2! 2!} \left\{ \begin{array}{cccc|cccc} a & b & c & d & \alpha & \beta & \gamma & \delta \\ e & f & g & h & \epsilon & \phi & \kappa & \\ i & & & & \lambda & & & \end{array} \right\}.$$

Any polynomial identity existing between e_1, e_2, \dots, e_m immediately implies a linear relation between the functions $e_{(\lambda)}$, say

$$\sum_{(\lambda)} c_{(\lambda)} e_{(\lambda)} = 0. \quad (7)$$

Double standard forms constructed from sets of independent variables are known to be linearly independent, and this leads one to guess that the $c_{(\lambda)}$ in the last equation are all zero. A little care, however, is always necessary in passing in this way from independent variables to equivalent symbols, as the principle of interchange of pairs of symbols often introduces linear relations between quantities which are otherwise independent. The vanishing of the $c_{(\lambda)}$ may be proved as follows. By the principle of interchange of equivalent symbols, each $e_{(\lambda)}$ in (7) may be replaced by $\phi_{(\lambda)}$, which is defined as $(1/n!)$ times the sum of all the double forms obtained from $\{S_{(\alpha)}^{(\lambda)} \mid S_{(\beta)}^{(\lambda)}\}$ by permuting the pairs of symbols $(a, \alpha), (b, \beta), \dots$ in all possible ways. Equation (7) now becomes

$$\sum c_{(\lambda)} \phi_{(\lambda)} = 0, \quad (8)$$

and here the symbols $a, \beta, \dots, a, b, \dots$ may be treated as independent variables, the symmetry of the $\phi_{(\lambda)}$ with respect to the different pairs of symbols replacing the principle of interchange of pairs of symbols. The vanishing of the $c_{(\lambda)}$ now follows by induction. In fact, a tableau of shape (λ) is said to precede a tableau of shape (μ) if the first row of the former which is not as long as the corresponding row of the latter is shorter than it. The hypothesis of the induction is that $c_{(\mu)} = 0$ for all shapes (μ) following the shape (λ) . But $c_{(\mu)}$ is zero if (μ) is the latest possible shape, namely the one-rowed tableau, as may be seen by putting $a=b=c=\dots$ and $\alpha=\beta=\gamma=\dots$; for this causes all the functions $\phi_{(\lambda)}$ to vanish except that corresponding to the single-rowed tableau. Then if any one of the

double forms in the expression $\phi_{(\lambda)}$ is selected, and all the symbols in each row separately of its tableaux are made equal to one another (so that, for example, $\left\{ \begin{array}{ccc|ccc} a & b & c & a & \beta & \gamma \\ d & e & & \delta & \epsilon & \end{array} \right\}$ becomes $\left\{ \begin{array}{ccc|ccc} a & a & a & a & a & a \\ b & b & & \beta & \beta & \beta \end{array} \right\}$), a non-vanishing double form will be obtained; but all the $\phi_{(\mu)}$ for shapes (μ) earlier than (λ) will be made to vanish by this coalescence of symbols, while the other forms appearing in $\phi_{(\mu)}$ will either vanish or become equal in value to the selected form. And so equation (8) implies that $c_{(\lambda)} = 0$, which completes the induction.

Hence the functions e_1, e_2, \dots, e_m form an irreducible rational integral basis for symmetric functions of the latent roots of A . And on account of the mutual expressibility (using equations (5) for $r=1, 2, \dots, m$) of the set e_1, e_2, \dots, e_m and s_1, s_2, \dots, s_m , the latter set is also an irreducible polynomial basis for symmetric functions.

The result just proved is, of course, a well-known one, but the method of proof is worth noting, for it illustrates a general method of reducing problems in equivalent symbols to problems in the manipulation of sets of independent variables. The result itself illustrates the second fundamental theorem of invariants, which in this case states that, apart from the identical vanishing of the M_r for $r > m$, any identity in symmetric functions depends entirely on algebraic manipulation of the inner products $a_\alpha, a_\beta, \dots, b_\alpha, b_\beta, \dots$. And so, on the basis of this theorem, the vanishing of the $c_{(\lambda)}$ in (7) follows at once. This method of proving the vanishing of the $c_{(\lambda)}$ as a consequence of the second fundamental theorem of invariants is, however, logically unsatisfactory, as the proof of this theorem is a matter of some difficulty.

The relations between the e_i and the s_i given by equation (5) are nothing else than Newton's relations between sums of powers and elementary symmetric functions; thus these relations, and also the classical fundamental theorem of symmetric functions, which states that the elementary symmetric functions form an irreducible polynomial basis for symmetric functions, have been derived by symbolic methods, using symbolic determinant reduction and the first fundamental theorem of projective invariants. There is thus a close link between the symbolic theory of determinants and that of the elementary symmetric functions e_i together with the s_i . This suggests that a corresponding link may subsist between the symbolic theory and that of the third well-known set of basic symmetric functions, namely $h_r \equiv N_r/r!$. The h_r have been shown to be representable as symbolic permanents; so the reduction of such a permanent will now be considered.

The working follows exactly the lines of the derivation of the Newton

relations, except that changes of sign are not involved when rows or columns of a permanent are interchanged. Starting with the symbolic permanent

$$(acde \dots || \beta\gamma\delta\epsilon \dots) \equiv \begin{vmatrix} a_\beta & a_\gamma & a_\delta & \dots \\ c_\beta & c_\gamma & c_\delta & \dots \\ d_\beta & d_\gamma & d_\delta & \dots \\ \vdots & \vdots & \vdots & \ddots \\ \vdots & \vdots & \vdots & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix}$$

of r rows and r columns, a reduction formula corresponding to (6) is obtained, namely

$$(acde \dots || \beta\gamma\delta\epsilon \dots) = a_\beta(cde \dots || \gamma\delta\epsilon \dots) + (r-1)c_\beta(acde \dots || \gamma\delta\epsilon \dots).$$

Then by repeated application of this reduction formula,

$$\begin{aligned} N_r &= (bcde \dots || \beta\gamma\delta\epsilon \dots) \\ &= b_\beta(cde \dots || \gamma\delta\epsilon \dots) + (r-1)c_\beta(bde \dots || \gamma\delta\epsilon \dots) \\ &= s_1 N_{r-1} + (r-1)c_\beta(bde \dots || \gamma\delta\epsilon \dots) \\ &= s_1 N_{r-1} + (r-1)s_2 N_{r-2} + (r-1)(r-2)c_\beta d_\gamma (de \dots || \delta\epsilon \dots) \\ &= s_1 N_{r-1} + (r-1)s_2 N_{r-2} + (r-1)(r-2)s_3 N_{r-3} + \dots + (r-1)! s_r. \end{aligned}$$

Hence

$$\frac{1}{(r-1)!} N_r = \frac{1}{(r-1)!} s_1 N_{r-1} + \frac{1}{(r-2)!} s_2 N_{r-2} + \dots + s_r,$$

i.e.

$$r h_r = s_1 h_{r-1} + s_2 h_{r-2} + \dots + s_r.$$

And these are Brioschi's relations between the sums of powers and the complete homogeneous functions h_r .

Newton's and Brioschi's relations may be solved, giving expressions for the s_i in terms of the e_i and h_i , respectively, or in terms of the M_i and N_i , respectively. The latter expressions are

$$\begin{aligned} s_1 &= M_1 &= N_1, \\ s_2 &= -M_2 + M_1^2 &= N_2 - N_1^2, \\ s_3 &= \frac{1}{2!}(M_3 - 3M_1 M_2 + 2M_1^3) &= \frac{1}{2!}(N_3 - 3N_1 N_2 + 2N_1^3), \text{ etc.} \end{aligned}$$

It may be verified by induction on i , after differentiating the i th Newtonian or Brioschian relation with regard to the first i of the

M_i and N_i , respectively, and adding the results, that

$$\sum_{j=1}^m j M_{j-1} \frac{\partial s_i}{\partial M_j} = 0 \quad \text{and} \quad \sum_{j=1}^m j N_{j-1} \frac{\partial s_i}{\partial N_j} = 0 \quad (i > 1). \quad (9)$$

Then, since the M_i appear as coefficients in the characteristic function of A , namely

$$f(\lambda) = |\lambda I - A| = \lambda^m - M_1 \lambda^{m-1} + \frac{M_2}{2!} \lambda^{m-2} \dots, \quad (10)$$

and since also

$$\frac{1}{f(\lambda)} = \frac{1}{\lambda^m} + \frac{N_1}{\lambda^{m+1}} + \frac{N_2}{2! \lambda^{m+2}} + \dots, \quad (11)$$

it follows that the s_i are seminvariants (of equal degree and weight) belonging to the binary *mic* $(1, M_1, M_2, \dots, M_m) \chi (X_1, X_2)^m$ and the binary *perpetuant* $(1, N_1, N_2, \dots) \chi (X_1, X_2)^m$ separately.

Incidentally, it is worth remarking that the differential equations (9) may be regarded as the annihilation of the s_i by each of the Sylvester operators

$$\Omega = \sum_j j M_{j-1} \frac{\partial}{\partial M_j}, \quad \Omega' = \sum_j j N_{j-1} \frac{\partial}{\partial N_j},$$

or equally well by the Hammond operators

$$d_1 = \sum_j e_{j-1} \frac{\partial}{\partial e_j}, \quad d'_1 = \sum_j h_{j-1} \frac{\partial}{\partial h_j},$$

(Hammond, 1882; also *cf.* MacMahon, 1915, I, 27). The polynomial expressions for the s_i in terms of the e_i and h_i are called *protomorphic seminvariants* (*cf.* Elliott, 1913, p. 212). Alternative solutions of the differential equations (9) exist. For instance, instead of s_4 may be taken

$$8e_4 - 8e_1e_3 + 4e_1^2e_2 - e_1^4 - s_2^2 - 2s_4,$$

with a similar expression in the h_i , giving $s_2^2 + 2s_4$. This illustrates the fact that complete systems of protomorphs, such as the s_i , are not necessarily unique.

One further result of the classical theory of symmetric functions remains to be discussed, namely the Wronskian relations between the e_i and the h_i which arise by multiplying together the series expansions (8) and (9) of $f(\lambda)$ and $1/f(\lambda)$ and equating the coefficients of powers of λ to zero. These relations, written in terms of the M_i and the N_i , are

$$\begin{aligned} M_1 - N_1 &= 0, \\ M_2 - 2M_1N_1 + N_2 &= 0, \\ M_3 - 3M_2N_1 + 3M_1N_2 - N_3 &= 0, \end{aligned}$$

and in general, say,

$$(M - N)_r \equiv M_r - rM_{r-1}N_1 + \frac{r(r-1)}{1.2}M_{r-2}N_2 - \dots + (-)^r N_r = 0,$$

with binomial coefficients (the corresponding relations between the ϵ_i and the h_i have unit coefficients). Writing these symbolically,

$$\begin{aligned}(a | a) - (a || a) &= 0, \\ (ab | a\beta) - 2(a | a)(a || a) + (ab || a\beta) &= 0, \\ (abc | a\beta\gamma) - 3(ab | a\beta)(c || \gamma) + 3(a | a)(bc || \beta\gamma) - (abc || a\beta\gamma) &= 0, \text{ etc.}\end{aligned}$$

But these are the fundamental identities between determinants and permanents discovered by Muir (Muir, 1897). In fact, we may write any one of them, say the third, in the more usual form,

$$|a_\alpha b_\beta c_\gamma| - \sum |a_\alpha b_\beta| |c_\gamma| + \sum |a_\alpha| |b_\beta c_\gamma| - |a_\alpha b_\beta c_\gamma| = 0,$$

where the summation is taken over all the different combinations of the pairs of symbols (a, α) , (b, β) , (c, γ) . The parallelism between Muir's results and the relations of Wronski has been clearly brought out by MacMahon (MacMahon, 1924; also MacMahon, 1927, p. 279; Littlewood, 1940, pp. 119-120), but the symbolic method shows it to be inevitable

7. SYMBOLIC EXPRESSION OF h -BIALTERNANTS

The symbolic double standard form $\{S_{(a)}^{(\lambda)} | S_{(a)}^{(\lambda)}\}$ has already been mentioned (§ 6). The polarized double standard form $\{\underline{S}_{(a)}^{(\lambda)} | S_{(a)}^{(\lambda)}\}$ is defined as the sum of all the double forms obtained by permuting in all possible ways the symbols within each row of $S_{(a)}^{(\lambda)}$ independently and adding up the results. Since $\{\underline{S}_{(a)}^{(\lambda)} | S_{(a)}^{(\lambda)}\}$ is obtained from $a_\alpha b_\beta c_\gamma \dots$ by operating on the sequence a, b, c, \dots with a certain substitutional operator, namely the Young operator E of the tableau $S_{(a)}^{(\lambda)}$ (E is one of the $E_r^{(\lambda)}$ of Rutherford (1948, p. 16)), this symbolic polarized double form is a symmetric function of the latent roots of A .

The symmetric functions $\{\underline{S}_{(a)}^{(\lambda)} | S_{(a)}^{(\lambda)}\}$ are of importance in the theory of invariant matrices, being, in fact, equal to numerical multiples of the traces of the irreducible invariant matrices of A . It may be quite simply shown from considerations of the structure of these invariant matrices that

$$\frac{1}{\theta^{(\lambda)}} \{\underline{S}_{(a)}^{(\lambda)} | S_{(a)}^{(\lambda)}\} = h_{(\lambda)}, \quad (12)$$

where $h_{(\lambda)}$ is the h -bialternant

$$\begin{vmatrix} & & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & h_{\lambda_1} & h_{\lambda_1+1} & h_{\lambda_1+2} \\ \cdot & \cdot & \cdot & h_{\lambda_1-1} & h_{\lambda_1} & h_{\lambda_1+1} \\ \cdot & \cdot & \cdot & h_{\lambda_1-2} & h_{\lambda_1-1} & h_{\lambda_1} \end{vmatrix}$$

in the latent roots of A ; $\theta^{(\lambda)}$ is a positive integer which may be defined by the equation $E^{\theta^{(\lambda)}} = \theta^{(\lambda)} E$ satisfied by the Young operator (Rutherford, 1948, p. 19).

Equation (12) may also be obtained in the following way, as was shown by Dr D. E. Rutherford in 1946.*

In the equation

$$\{S_{(a)}^{(\lambda)} | S_{(a)}^{(\lambda)}\} = E a_a b_b c_c \dots$$

on account of the interchangeability property of equivalent symbols, E may be replaced by $\sigma E \sigma^{-1}$, where σ is any permutation belonging to the symmetric group on the n letters a, b, c, \dots . But $\sigma E \sigma^{-1}$ is the Young operator corresponding to a tableau of shape (λ) in the a, b, c, \dots obtained by performing the permutation σ on the symbols of $S_{(a)}^{(\lambda)}$. And so

$$\begin{aligned} \{S_{(a)}^{(\lambda)} | S_{(a)}^{(\lambda)}\} &= E a_a b_b c_c \dots \\ &= \frac{1}{n!} \sum_{\sigma} \sigma E \sigma^{-1} a_a b_b c_c \dots \end{aligned}$$

(the summation being over all elements of the symmetric group on n letters)

$$= \frac{(\theta^{(\lambda)})^2}{n!} T^{(\lambda)} a_a b_b c_c \dots$$

(Rutherford, 1948, p. 65). Here $T^{(\lambda)}$ is a substitutional operator, introduced by Young, and defined as

$$\sum_{i=1}^{f^{(\lambda)}} g_{ii}^{(\lambda)},$$

where the $n!$ substitutional operators $g_{ii}^{(\lambda)}$ are basal units for any set of irreducible representations of the symmetric group and $f^{(\lambda)}$ is the number of standard tableaux of shape (λ) in n distinct symbols. From the theory

* Turnbull communicated this formula, which he had verified for simple cases by elementary methods, to the Edinburgh Mathematical Society in the session 1945-46, and at a seminar at St Andrews, when he asked for a formal proof. The above proof was supplied next day in a letter by Dr D. E. Rutherford.

of group characters (Rutherford, 1948, p. 66),

$$T^{(\lambda)} = \frac{1}{\theta^{(\lambda)}} \sum_{(\rho)} \chi_{(\rho)}^{(\lambda)} C_{(\rho)},$$

where $\chi_{(\rho)}^{(\lambda)}$ is the character of a permutation of class (ρ) in the representation corresponding to partition (λ) . And so, since $f^{(\lambda)}\theta^{(\lambda)} = n!$ (Rutherford, 1948, p. 65),

$$\begin{aligned} \{S_{(a)}^{(\lambda)} | S_{(a)}^{(\lambda)}\} &= \frac{1}{f^{(\lambda)}} \sum_{(\rho)} \chi_{(\rho)}^{(\lambda)} C_{(\rho)} a_a b_{\rho} c_{\gamma} \dots \\ &= \frac{1}{f^{(\lambda)}} \sum_{(\rho)} \chi_{(\rho)}^{(\lambda)} k_{(\rho)} S_{(\rho)} \quad (\S 4, \text{equation (3)}). \end{aligned}$$

But since, from the Frobenius theory of group characters,

$$n! h_{(\lambda)} = \sum_{(\rho)} \chi_{(\rho)}^{(\lambda)} k_{(\rho)} S_{(\rho)}$$

(Littlewood, 1940, p. 86), equation (12) follows. The result is implicit also in Theorem II of Young's Fourth Memoir (Young, 1929, p. 259).

A simple example of (12) involving three pairs of equivalent symbols is

$$\frac{1}{3} \left[\left\{ \begin{array}{cc|cc} a & b & a & \beta \\ c & & \gamma & \end{array} \right\} + \left\{ \begin{array}{cc|cc} b & a & a & \beta \\ c & & \gamma & \end{array} \right\} \right] = \left| \begin{array}{cc} h_2 & h_3 \\ h_0 & h_1 \end{array} \right|.$$

The expanded form on the left is $\frac{1}{3}(a_a b_{\rho} c_{\gamma} - a_{\rho} b_{\gamma} c_a)$, which is $\frac{1}{3}(s_1^3 - s_3)$, which again is equal to the h -bialternant on the right, and indeed also to $e_2 e_1 - e_0 e_3$.

REFERENCES TO LITERATURE

- ARONHOLD, 1858. "Theorie der homogenen Funktionen dritten Grades von drei Veränderlichen", *Journ. für Math.*, LV, 97-191.
 CLEBSCH, 1861 *a*. "Über eine Transformation der homogenen Funktionen dritter Ordnung mit vier Veränderlichen", *Journ. für Math.*, LVIII, 109-126.
 —, 1861 *b*. "Über symbolische Darstellung algebraischer Formen", *Journ. für Math.*, LIX, 1-62.
 ELLIOTT, 1913. *Algebra of Quantics*, Oxford.
 HAMMOND, 1882. "On the Calculating of Symmetric Functions", *Proc. London Math. Soc.*, ser. 1, XIII, 79.
 LITTLEWOOD, 1940. *The Theory of Group Characters*, Oxford.
 MACMAHON, 1915. *Combinatory Analysis*, Cambridge.
 —, 1924. "Researches in the Theory of Determinants", *Trans. Cambridge Phil. Soc.*, XXIII, 89-135.

- MACMAHON, 1927. "The Structure of a Determinant", *Journ. London Math. Soc.*, II, 273-286.
- MUIR, 1897. "A Relation between Permanents and Determinants", *Proc. Roy. Soc. Edin.*, XXII, 134-136.
- RUTHERFORD, 1948. *Substitutional Analysis*, Edinburgh.
- SYLVESTER, 1851. "On the Relation between the Minor Determinants of Linearly Equivalent Quadratic Functions", *Phil. Mag.*, ser. 4, I, 295-305, 415; *Coll. Math. Papers*, I, 241-250, 251.
- , 1852. "On the Principles of the Calculus of Forms", *Cambridge and Dublin Math. Journ.*, VII, 52-97; *Coll. Math. Papers*, I, 284-327.
- TURNBULL, 1932. "The Invariant Theory of a General Bilinear Form", *Proc. London Math. Soc.*, ser. 2, XXXIII, 1-19.
- , 1945. *The Theory of Determinants, Matrices and Invariants* (2nd Ed.), London and Glasgow.
- YOUNG, 1929. "On Quantitative Substitutional Analysis" (Fourth Memoir), *Proc. London Math. Soc.*, ser. 2, XXXI, 253-288.

(Issued separately May 14, 1951)

XIII.—Studies in Practical Mathematics. VI. On the Factorization of Polynomials by Iterative Methods.* By **A. C. Aitken**, D.Sc., F.R.S., Mathematical Institute, University of Edinburgh.

(MS. received March 27, 1950. Revised MS. received July 31, 1950. Read June 5, 1950)

SYNOPSIS

The method of iteration of penultimate remainders, introduced by S. N. Lin for approximating by stages to the exact factors of a polynomial, is subjected to theoretical analysis. The matrix governing the iterative process is obtained, and its latent roots and latent vectors are found. Incidental theorems yielding further factorizations are proved, and processes are developed for accelerating convergence. Numerical examples illustrate varying situations likely to arise in practice.

1 PRELIMINARY

THE purpose of this paper is to supply the necessary theory and to extend the scope of an iterative method (Lin, 1941) for approximating by stages to an exact factor of a polynomial. Other writers (Fry, 1945; Friedman, 1949) have discussed the method. Friedman has introduced a procedure which he describes as a modification of Lin's; but in fact, though both methods make use of iterated polynomial division, Friedman's method is his own, and rests on a different basis from that of Lin.

We examine the process *ab initio*, and extend it to divisors of arbitrary degree. We regard it from the point of view of repeated linear transformation of a vector, namely the vector of small errors or deviations from the final exact coefficients. The linear operator that emerges is a matrix R of special type; expressions are obtained for its latent roots and its latent vectors. Not only the dominant latent root, on which the convergence of the process depends, but the corresponding latent vector reveal themselves in the course of the arithmetic and provide (§ 7) valuable additional information. It will appear also that certain processes for accelerating convergence, familiar already in other applications of vector iteration, are again available here; and we introduce a new and simplified version of one of these, adapted to the case where the dominant latent roots of R are a conjugate complex pair.

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

2. THE PENULTIMATE REMAINDER POLYNOMIAL

Let the polynomial $f_n(x)$ proposed for factorization be

$$f_n(x) = x^n - a_1x^{n-1} + a_2x^{n-2} - \dots + (-)^na_n, \quad a_n \neq 0, \quad (1)$$

$$= (x - a_1)(x - a_2) \dots (x - a_n), \quad a_1a_2 \dots a_n \neq 0, \quad (2)$$

where the divisor polynomial (we shall later specify which particular divisor)

$$d_m(x) = x^m - b_1x^{m-1} + b_2x^{m-2} - \dots + (-)^mb_m \quad (3)$$

$$= (x - \beta_1)(x - \beta_2) \dots (x - \beta_m), \quad (4)$$

and the corresponding quotient polynomial for this divisor is

$$q_{n-m}(x) = x^{n-m} - c_1x^{n-m-1} + c_2x^{n-m-2} - \dots + (-)^{n-m}c_{n-m} \quad (5)$$

$$= (x - \gamma_1)(x - \gamma_2) \dots (x - \gamma_{n-m}). \quad (6)$$

The above is a resolution into exact polynomial factors. In practice, however, the factors of $f_n(x)$ are not known, and the problem is to obtain them. For this purpose we begin by taking a trial divisor of degree m , let us say

$$t_m(x) = x^m - b'_1x^{m-1} + b'_2x^{m-2} - \dots + (-)^mb'_m, \quad (7)$$

presumed to be an approximation to some exact divisor $d_m(x)$. Using the routine of polynomial division, in a form adapted for machine (§ 5) if one is available, we divide $f_n(x)$ by $t_m(x)$. The division is stopped not, as usual, at the final remainder, but at the stage immediately preceding, where the current remainder is of the same degree, m , as the divisor. This particular remainder will be called the *penultimate remainder* (p.r.) of $f_n(x)$ with respect to $t_m(x)$. When divided by its leading coefficient, so that the highest term becomes x^m , the p.r. yields the *reduced penultimate remainder* (r.p.r.).

The method now to be investigated is that of *iteration of the operation of penultimate remaindering*, each successive r.p.r. being taken as a fresh divisor. It will appear that under certain conditions the sequence of r.p.r. polynomials converges to an exact factor $d_m(x)$ of $f_n(x)$. There are in general $\binom{n}{m}$ such exact factors of degree m , real or complex; each is characterized (§ 8) by stability or instability, in the sense that a slightly displaced divisor yields an r.p.r. in the one case nearer to $d_m(x)$, in the other case further away from it. The criteria of stability (§ 4) depend both on the divisor $d_m(x)$ and on the quotient $q_{n-m}(x)$.

The coefficients both in quotient and in p.r. are determinants of a

special bigradient, almost persymmetric recurrent, type not unfamiliar in the literature. When $f_n(x)$ is divided by $t_m(x)$ the quotient is as follows:

$$x^{n-m} + \begin{vmatrix} 1 & a_1 \\ 1 & b_1' \end{vmatrix} x^{n-m-1} + \begin{vmatrix} 1 & a_1 & a_2 \\ 1 & b_1' & b_2' \\ 1 & & b_1' \end{vmatrix} x^{n-m-2} + \dots; \quad (8)$$

while the form of the p.r. may be sufficiently indicated by the first few terms in the case $n-m=3$, namely

$$\begin{vmatrix} 1 & a_1 & a_2 & a_3 \\ 1 & b_1' & b_2' & b_3' \\ & 1 & b_1' & b_2' \\ & & 1 & b_1' \end{vmatrix} x^m + \begin{vmatrix} 1 & a_1 & a_2 & a_4 \\ 1 & b_1' & b_2' & b_4' \\ & 1 & b_1' & b_3' \\ & & 1 & b_2' \end{vmatrix} x^{m-1} + \begin{vmatrix} 1 & a_1 & a_2 & a_5 \\ 1 & b_1' & b_2' & b_5' \\ & 1 & b_1' & b_4' \\ & & 1 & b_3' \end{vmatrix} x^{m-2} + \dots \quad (9)$$

Dividing the p.r. by its leading coefficient, presumed not to vanish, we have the r.p.r.

3. THE LINEAR OPERATION ON THE VECTOR OF ERRORS

Let $b_i' = b_i + \epsilon_i$, and let it be assumed that, relative to the moduli of the coefficients b_i , the moduli of the errors $\epsilon_i, \epsilon_j, \dots$ are so small that powers and products above the first degree may be neglected. We proceed to study, under this assumption, the transformation of errors produced by the p.r. operation. It will appear that this transformation is represented by a matrix R of m rows and m columns. The elements of R are determinants of varying order. To write R out *in extenso* for general n and m would require a great deal of space; but the fundamental results and their proof can be sufficiently illustrated by the case $n=7, m=4, n-m=3$.

Let us first examine, in this case, the coefficient of x^{m-1} in the r.p.r., that is to say, the coefficient corresponding to b_1 in $d_m(x)$. It is

$$b_1'' = \begin{vmatrix} 1 & a_1 & a_2 & a_4 \\ 1 & b_1 + \epsilon_1 & b_2 + \epsilon_2 & b_4 + \epsilon_4 \\ & 1 & b_1 + \epsilon_1 & b_3 + \epsilon_3 \\ & & 1 & b_2 + \epsilon_2 \end{vmatrix} \div \begin{vmatrix} 1 & a_1 & a_2 & a_3 \\ 1 & b_1 + \epsilon_1 & b_2 + \epsilon_2 & b_3 + \epsilon_3 \\ & 1 & b_1 + \epsilon_1 & b_2 + \epsilon_2 \\ & & 1 & b_1 + \epsilon_1 \end{vmatrix}. \quad (1)$$

The deviation $\epsilon_1^{(1)} = b_1'' - b_1$ is readily found to be

$$\epsilon_1^{(1)} = - \begin{vmatrix} 1 & a_1 & a_2 & a_3 & a_4 \\ 1 & b_1 + \epsilon_1 & b_2 + \epsilon_2 & b_3 + \epsilon_3 & b_4 + \epsilon_4 \\ & 1 & b_1 + \epsilon_1 & b_2 + \epsilon_2 & b_3 + \epsilon_3 \\ & & 1 & b_1 + \epsilon_1 & b_2 + \epsilon_2 \\ & & & 1 & b_1 \end{vmatrix} \div \begin{vmatrix} 1 & a_1 & a_2 & a_3 \\ 1 & b_1 + \epsilon_1 & b_2 + \epsilon_2 & b_3 + \epsilon_3 \\ & 1 & b_1 + \epsilon_1 & b_2 + \epsilon_2 \\ & & 1 & b_1 + \epsilon_1 \end{vmatrix}. \quad (2)$$

But, comparing coefficients of powers x^r in § 2 (1) and the product of § 2 (3) and § 2 (5), we have

$$a_r = b_r + c_1 b_{r-1} + c_2 b_{r-2} + \dots + c_{r-1} b_1 + c_r. \quad (3)$$

Hence, performing the operation

$$\text{row}_1 - \text{row}_2 - c_1 \text{row}_2 - c_2 \text{row}_4 - c_3 \text{row}_5 \quad (4)$$

on the numerator in (2), with a similar operation on the denominator, and retaining terms of the first order only in the ϵ , we have on expansion

$$\epsilon_1^{(1)} = (-)^{s-1} c_s^{-1} \left\{ \begin{vmatrix} 1 & c_1 & c_2 & . \\ 1 & b_1 & b_2 & b_3 \\ & 1 & b_1 & b_2 \\ & & 1 & b_1 \end{vmatrix} \epsilon_1 - \begin{vmatrix} 1 & c_1 & c_2 \\ 1 & b_1 & b_2 \\ & 1 & b_1 \end{vmatrix} \epsilon_2 + \begin{vmatrix} 1 & c_1 \\ 1 & b_1 \end{vmatrix} \epsilon_3 - \epsilon_4 \right\}. \quad (5)$$

An analogous result holds for general values of n and m , and can be proved in a similar manner. A similar procedure can be applied to the determinant-quotients in the coefficients of the r.p.r. corresponding to b_2 , b_3 and b_4 . The results can be summed up in the matrix operation on the *vector* of errors ϵ , namely

$$\epsilon^{(1)} = R\epsilon, \quad (6)$$

where R in our illustration is the matrix

$$R = (-)^{s-1} c_s^{-1} \begin{bmatrix} \begin{vmatrix} 1 & c_1 & c_2 & . \\ 1 & b_1 & b_2 & b_3 \\ & 1 & b_1 & b_2 \\ & & 1 & b_1 \end{vmatrix} & - \begin{vmatrix} 1 & c_1 & c_2 \\ 1 & b_1 & b_2 \\ & 1 & b_1 \end{vmatrix} & \begin{vmatrix} 1 & c_1 \\ 1 & b_1 \end{vmatrix} & -1 \\ \begin{vmatrix} 1 & c_1 & c_2 & . \\ 1 & b_1 & b_2 & b_4 \\ & 1 & b_1 & b_3 \\ & & 1 & b_2 \end{vmatrix} & - \begin{vmatrix} 1 & c_1 & . \\ 1 & b_1 & b_3 \\ & 1 & b_2 \end{vmatrix} & \begin{vmatrix} 1 & c_2 \\ 1 & b_2 \end{vmatrix} & -c_1 \\ \begin{vmatrix} 1 & c_1 & c_2 & . \\ 1 & b_1 & b_2 & . \\ & 1 & b_1 & b_4 \\ & & 1 & b_3 \end{vmatrix} & - \begin{vmatrix} 1 & c_1 & . \\ 1 & b_1 & b_4 \\ & 1 & b_3 \end{vmatrix} & \begin{vmatrix} 1 & . \\ 1 & b_3 \end{vmatrix} & -c_2 \\ \begin{vmatrix} 1 & c_1 & c_2 & . \\ 1 & b_1 & b_2 & . \\ & 1 & b_1 & . \\ & & 1 & b_4 \end{vmatrix} & - \begin{vmatrix} 1 & c_1 & . \\ 1 & b_1 & . \\ & 1 & b_4 \end{vmatrix} & \begin{vmatrix} 1 & . \\ 1 & b_4 \end{vmatrix} & . \end{bmatrix}. \quad (7)$$

In the general case R is a matrix of order $m \times m$, and is derived by steps quite analogous to those described above. The general form of R can be inferred from inspection of (7). It is chiefly characterized by its leading element, the later elements in the first row being principal minors of this leading element. At each descent to a lower row the elements in the *last column* of the determinants concerned receive a unit increase of suffix. A feature worth noticing is that c_{n-m} , represented in the example above by c_3 , though present in the extraneous scalar factor $(-)^{n-m-1}c_{n-m}^{-1}$, is missing from the elements of the determinants within R .

Each successive r.p.r. is taken as the divisor for a new r.p.r.; and so the vector of errors, or more precisely the vector of the *linear* components of the errors, is repeatedly transformed by R . Hence the condition of convergence, under the stated assumptions, is that all the latent roots ρ_r of R should be such that $|\rho_r| < 1$. We must next obtain these latent roots and the associated latent vectors.

4. LATENT ROOTS AND VECTORS OF THE MATRIX OF ITERATION

The zeros β_i of $d_m(x)$ together with the zeros γ_i of $q_{n-m}(x)$ constitute all the n zeros of $f_n(x)$. It will be shown that in the general case the m latent vectors of R are

$$\left\{ 1 \quad \sum_{[v]} \beta_i \quad \sum_{[v]} \beta_i \beta_j \quad \sum_{[v]} \beta_i \beta_j \beta_k \dots \right\}, \quad (1)$$

where the bracketed subscript $[v]$ denotes that in the summations (each of which yields an elementary symmetric function of a particular set of $m-1$ roots β_i) one root β_v is omitted in each latent vector; or is replaced by 0, which has the same effect. Thus in the respective vectors we omit in turn $\beta_1, \beta_2, \dots, \beta_m$.

As for the latent roots ρ_r , they are of a rather special nature, most simply described as coefficients in *reduced penultimate quotients* (r.p.q.) with respect to linear divisors $x - \beta_v$. In fact, let the penultimate quotient when $f_n(x)$ is divided by $d_m(x)$ be reduced by dividing it through by $(-)^{n-m-1}c_{n-m}$, the constant term of the final quotient with sign changed. The result, which we shall call the r.p.q., is

$$(-)^{n-m-1}c_{n-m}^{-1}\{x^{n-m} - c_1x^{n-m-1} + c_2x^{n-m-2} - \dots + (-)^{n-m-1}c_{n-m-1}x\}. \quad (2)$$

We now assert that the latent roots ρ_r of R are the m values assumed by this r.p.q. when $x = \beta_1, \beta_2, \dots, \beta_m$. The determinantal expression for ρ_r , namely

$$\rho_r = (-)^{n-m-1} c_{n-m}^{-1} \begin{vmatrix} 1 & c_1 & c_2 & c_3 & \dots & c_{n-m-1} & \cdot \\ 1 & \beta_r & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & 1 & \beta_r & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & 1 & \beta_r & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \beta_r & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \beta_r & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \beta_r \end{vmatrix}, \quad (3)$$

shows how ρ_r is related to the leading element of R . It also shows that ρ_r is the coefficient of x in the r.p.q. arising when $q_{n-m}(x)$ is divided by $x - \beta_r$.

Deferring for the time being the proof of these statements, let the m latent roots ρ_r be named in such an order that

$$|\rho_1| > |\rho_2| > |\rho_3| > \dots > |\rho_m|. \quad (4)$$

It is hardly necessary to point out that this order does not usually correspond to the descending order of the $|\beta_r|$. Then the sufficient condition of convergence of the r.p.r. process is that $|\rho_1| < 1$.

In the not infrequent case where ρ_1 is real and $|\rho_1| > |\rho_2|$, convergence will be of the kind in which the successive errors of corresponding coefficients in the r.p.r.'s, or equally well the first differences of corresponding coefficients between consecutive r.p.r.'s, tend to *geometric progressions* of common ratio ρ_1 . The known accelerative processes based upon this (Aitken, 1926; Steffensen, 1933) are then available, and should be applied at as early a stage as possible.

If, as is also not unusual, ρ_1 is one of a conjugate complex pair, the familiar features of oscillatory convergence resembling a damped vibration will be present. Here again a simple accelerative process (§ 6) is available.

If $|\rho_1| > 1$, the sequence of r.p.r.'s will diverge. In some cases, but not all, convergence may be restored by the use of the reciprocal equation, that is, by performing penultimate remaindering with reversed order of terms. On the other hand it may be advantageous to give up trying a divisor m and to choose one of degree $m - 1$ or $m + 1$.

We proceed to establish the theorems enunciated above. As remarked earlier, to economize in space and to avoid undue prolixity, we illustrate by the case $n = 7$, $m = 4$. We prove first that, for example,

$$R \begin{bmatrix} 1 \\ \beta_2 + \beta_3 + \beta_4 \\ \beta_2\beta_3 + \beta_2\beta_4 + \beta_3\beta_4 \\ \beta_2\beta_3\beta_4 \end{bmatrix} = c_3^{-1} \begin{vmatrix} 1 & c_1 & c_2 & \cdot \\ 1 & \beta_1 & \cdot & \cdot \\ \cdot & 1 & \beta_1 & \cdot \\ \cdot & \cdot & 1 & \beta_1 \end{vmatrix} \begin{bmatrix} 1 \\ \beta_2 + \beta_3 + \beta_4 \\ \beta_2\beta_3 + \beta_2\beta_4 + \beta_3\beta_4 \\ \beta_2\beta_3\beta_4 \end{bmatrix}, \quad (5)$$

where R is as given in § 3 (7); with three results of similar form. The latent root and latent vector are visible in the above equation.

To prove this result we first refer to § 2 (8), which shows that, apart from sign and reversed order, the elements in the first row of R (see again § 3 (7)) are the respective coefficients in successive terms of the quotient resulting from the *formal* division of the r.p.q. $c_3^{-1}(x^3 - c_1x^2 + c_2x)$ by $d_m(x)$. Let this r.p.q. be multiplied by

$$(x - \beta_1)(x - \beta_2)(x - \beta_3) = x^3 - (\beta_1 + \beta_2 + \beta_3)x^2 + (\beta_1\beta_2 + \beta_1\beta_3 + \beta_2\beta_3)x - \beta_1\beta_2\beta_3,$$

and let us suppose the resulting product to be divided by $d_m(x)$. The coefficient in the fourth term of the quotient so obtained will then be the linear combination

$$c_3^{-1} \left[\begin{vmatrix} 1 & c_1 & c_2 & . \\ 1 & \beta_1 & \beta_2 & \beta_3 \\ & 1 & \beta_1 & \beta_2 \\ & & 1 & \beta_1 \end{vmatrix} - \begin{vmatrix} 1 & c_1 & c_2 \\ 1 & \beta_1 & \beta_2 \\ & 1 & \beta_1 \end{vmatrix} \begin{vmatrix} 1 & c_1 \\ 1 & \beta_1 \end{vmatrix} - 1 \right] \\ \{1 \quad \beta_1 + \beta_2 + \beta_3 \quad \beta_1\beta_2 + \beta_1\beta_3 + \beta_2\beta_3 \quad \beta_1\beta_2\beta_3\}. \quad (6)$$

But since

$$d_m(x) = (x - \beta_1)\{x^3 - (\beta_2 + \beta_3 + \beta_4)x^2 + (\beta_2\beta_3 + \beta_2\beta_4 + \beta_3\beta_4)x - \beta_2\beta_3\beta_4\}, \quad (7)$$

it follows that (6) will be the coefficient in the fourth term of the quotient arising from the division of the r.p.q. $c_3^{-1}(x^3 - c_1x^2 + c_2x)$ by $x - \beta_1$; and so, again by reference to § 2 (8), it will be

$$c_3^{-1} \begin{vmatrix} 1 & c_1 & c_2 & . \\ 1 & \beta_1 & . & . \\ & 1 & \beta_1 & . \\ & & 1 & \beta_1 \end{vmatrix} = c_3^{-1}(\beta_1^3 - c_1\beta_1^2 + c_2\beta_1). \quad (8)$$

Thus (5) is established, as far as the first row of R is concerned. As for the second row of R , we shall indicate determinants by their diagonal elements (with the convention that $\delta_0 = c_0 = 1$) and we shall write it in the form

$$c_1^{-1} \left[- |c_0\delta_1\delta_1\delta_1| + \delta_1 |c_0\delta_1\delta_1\delta_1| \quad |c_0\delta_1\delta_1\delta_1| - \delta_1 |c_0\delta_1\delta_1| \right. \\ \left. - |c_0\delta_1\delta_1| + \delta_1 |c_0\delta_1| \quad |c_0\delta_1| - \delta_1 \right]. \quad (9)$$

Applying the same reasoning as in (6), (7) and (8), we find that the second

element in $R\{\begin{smallmatrix} 1 & \beta_2 + \beta_3 + \beta_4 & \beta_2\beta_3 + \beta_2\beta_4 + \beta_3\beta_4 & \beta_2\beta_3\beta_4 \end{smallmatrix}\}$ is

$$-c_3^{-1} \begin{vmatrix} 1 & c_1 & c_2 & . & . \\ 1 & \beta_1 & . & . & . \\ & 1 & \beta_1 & . & . \\ & & 1 & \beta_1 & . \\ & & & 1 & \beta_1 \end{vmatrix} + c_3^{-1} b_1 \begin{vmatrix} 1 & c_1 & c_2 & . \\ 1 & \beta_1 & . & . \\ & 1 & \beta_1 & . \\ & & 1 & \beta_1 \end{vmatrix} \quad (10)$$

$$= c_3^{-1} (\beta_2 + \beta_3 + \beta_4) (\beta_1^3 - c_1 \beta_1^2 + c_2 \beta_1), \quad (11)$$

since we have

$$b_1 - \beta_1 = \beta_2 + \beta_3 + \beta_4. \quad (12)$$

In similar fashion the third row of R may be written

$$c_3^{-1} \left[- |c_0 b_1 b_1 b_1 b_1| + b_2 |c_0 b_1 b_1 b_1| \quad |c_0 b_1 b_1 b_1| - b_2 |c_0 b_1 b_1| \right. \\ \left. - |c_0 b_1 b_2| + b_2 |c_0 b_1| \quad |c_0 b_2| - b_2 \right], \quad (13)$$

and we find in the same way that the third element in

$$R\{\begin{smallmatrix} 1 & \beta_2 + \beta_3 + \beta_4 & \beta_2\beta_3 + \beta_2\beta_4 + \beta_3\beta_4 & \beta_2\beta_3\beta_4 \end{smallmatrix}\}$$

is

$$c_3^{-1} (\beta_2 \beta_3 + \beta_2 \beta_4 + \beta_3 \beta_4) (\beta_1^3 - c_1 \beta_1^2 + c_2 \beta_1), \quad (14)$$

since we have

$$b_2 - \beta_1 (\beta_2 + \beta_3 + \beta_4) = \beta_2 \beta_3 + \beta_2 \beta_4 + \beta_3 \beta_4. \quad (15)$$

The fourth element is determined in similar fashion. Thus (5) is established. The corresponding results for the remaining latent roots and latent vectors of R follow at once from symmetry. In the general case where R is of order $m \times m$ a proof can be constructed following the same lines.

5. APPLICATION TO NUMERICAL EXAMPLES

It is instructive at this point to work out some numerical examples, both with real and with complex dominant latent roots of R . So that the nature and rapidity of the convergence shall be more clearly visible, we have purposely chosen polynomials $f_n(x)$ having factors with integer coefficients.

Example 1.—To find the real root of Wallis's equation

$$f_2(x) = x^3 - 2x - 5 = 0.$$

By inspection the root is slightly greater than 2. So we transform by $x = s + 2$ and consider the equation

$$s^3 + 6s^2 + 10s - 1 = 0.$$

Note.—In polynomial division by machine a condensed form of synthetic division is most efficacious. Nothing need be written down on the computing sheet except the coefficients c_i of the quotient, obtained one after another, concluding with the final remainder. In our modified procedure we should similarly obtain the coefficients of the penultimate quotient and penultimate remainder. For example, giving positive sign to all coefficients so as to show procedure more simply, if we have to divide

$$x^n + a_1x^{n-1} + a_2x^{n-2} + \dots \text{ by } x^m + b_1x^{m-1} + b_2x^{m-2} + \dots, \quad (1)$$

we compute and write down in a row the numerical values of

$$1, \quad c_1 = a_1 - b_1, \quad c_2 = a_2 - c_1b_1 - b_2, \quad c_3 = a_3 - c_1b_2 - c_2b_1 - b_3, \quad \dots \quad (2)$$

the coefficients of the penultimate remainder automatically filling up the rest of the row. Then in a new row we enter the detached coefficients of the r.p.r., the new divisor, obtained from the p.r. by multiplying through by the reciprocal of the leading coefficient, then the new penultimate quotient obtained as in (2) and then the new p.r., and so on. The summing of the signed isobaric products in (2) might be aided by having the coefficients $1, b_1, b_2, \dots$ suitably spaced on a loose card which could be ranged above the appropriate c_i .

In the present example the computing sheet will read as follows:

r.p.r.	quotient	p.r.
1 -0.1	1 6	10 -1
1 -0.094	6.1	10.61 -1
1 -0.0945832	6.094	10.5728 -1
1 -0.0945497	6.0945832	10.576445 -1
1 -0.0945516	6.0945497	10.576238 -1
1 -0.0945515	6.0945516	10.576249 -1

For example, $10.576445 = 10 - (-0.0945832)(6.0945832)$. The p.r. in this row is thus $10.576445x - 1$; reduced by division by 10.576445 it gives the r.p.r. $x - 0.0945497$ for the next division.

The required root is $x = 0.0945515$, $x = 2.0945515$. We shall see in § 6 that with slight increase of work this could be improved by several additional digits.

Note.—It calls for remark that, so far as division by a *linear* factor is concerned, penultimate remaindering is really not new; it is merely a paraphrase of the old-established process (Whittaker and Robinson, 1929) of iteration. Iteration consists in writing an equation $\varphi(x) = 0$ in a suitable form $f(x) = F(x)$, and in building a sequence x_k satisfying the recurrence $f(x_k) = F(x_{k-1})$. In the present case the choice is

$$f(x) = x, \quad F(x) = (-)^{n-1}a_n\{x^{n-1} - a_1x^{n-2} + \dots + (-)^{n-1}a_{n-1}\}^{-1}, \quad (3)$$

and the initial x_1 is simply the b_1' of § 2. Penultimate remaindering is, step for step, equivalent to the following way of numerically evaluating the polynomial within the bracket on the right of (3):

$$x_k(x_k - a_1(x_k - a_2(x_k - \dots))) + (-)^{n-1}a_{n-1}. \quad (4)$$

This is of course a well-known way, on a machine provided with a transfer lever from product to setting register, for evaluating a polynomial by stages; it requires no intermediate copying.

Example 2.—We shall try to find a cubic factor of the polynomial

$$f_4(x) = x^4 + 17x^3 + 84x^2 + 148x + 80.$$

The very crude choice x^3 for first divisor would give the last four terms for p.r., roughly reduced to $x^3 + 5x^2 + 9x + 5$, and a trial with this again suggests $x^3 + 6x^2 + 12x + 7$ as a closer divisor. From that point the computing sheet reads as follows:

r.p.r.			q.	p.r.				
1	6.0	12.0	7.0	1	17.0	84.0	148.0	80.0
	6.54545	12.81818	7.27273		11.0	72.0	141.0	80.0
	6.80869	13.46086	7.65217		10.45455	71.18182	140.72727	...
	6.92150	13.77132	7.84982		10.19131	70.53914	140.34783	...
	6.96817	13.90586	7.93769		10.07850	70.22868	140.15018	
	6.98717	13.96179	7.97462		10.03183	70.09414	140.06231	
	6.99485	13.98460	7.98975		10.01283	70.03821	140.02538	

Here we see a fair convergence to the factor $x^3 + 7x^2 + 14x + 8$. The ratio of convergence is in fact 0.4. The r.p.q. is $-x/10$, and the zeros of the divisor are $-1, -2, -4$. In § 6 we shall use this example to illustrate the acceleration of convergence.

Example 3.—In the quartic polynomial

$$f_4(x) = x^4 + 16x^3 + 71x^2 + 122x + 120$$

the zeros of the divisor will prove to be complex. We shall try for a quadratic factor, beginning with $x^2 + 1.7x + 1.7$, suggested by inspection of the last three terms of the quartic. The sheet reads thus:

r.p.r.			q.	p.r.			
I	1.7	1.7	I	16.0	71.0	122.0	120.0
	2.17137	2.66726		14.3	44.99	97.69	120.0
	2.22201	3.13270		13.82863	38.30567	85.11545	...
	2.11631	3.22126		13.77799	37.25247	78.83769	
	2.01260	3.12528		13.88369	38.39655	77.27702	
	1.97075	3.02087		13.98740	39.72368	78.28546	
	1.97415	2.97538		14.02925	40.33099	79.61946	
	1.99000	2.97505		14.02585	40.33549	80.26777	
	2.00073	2.98916		14.01000	40.14505	80.31955	

The convergence towards the exact factor $x^2 + 2x + 3$ is clearly seen, as also is the oscillation of error produced by the presence of dominant conjugate complex roots of the governing matrix R . We shall also use this example in § 6 to illustrate acceleration of convergence.

Example 4.—Finally, in order to show the emergence of a valuable feature not hitherto mentioned, we shall take the same polynomial as in Example 3, and shall try for a cubic divisor, beginning with $x^3 + 4x^2 + 8x + 8$. The sheet reads:

r.p.r.			q.	p.r.				
I	4.0	8.0	8.0	I	16.0	71.0	122.0	120
	5.25	9.5	10.0		12.0	63.0	114.0	120
	5.72093	10.41860	11.16279		10.75	61.5	112.0	...
	5.89367	10.78281	11.67421		10.27907	60.58140	110.83721	
	5.95836	10.91650	11.87375		10.10633	60.21719	110.32579	
	5.98343	10.96696	11.95024		10.04164	60.08350	110.12625	
	5.99337	10.98677	11.98015		10.01657	60.03304	110.04976	
	5.99735	10.99470	11.99205		10.00663	60.01323	110.01985	

Here we see a convergence towards the exact factor $x^3 + 6x^2 + 11x + 12$. This factor has a pair of complex zeros, but they are not dominant in the process. The dominant one is $x = -4$, the r.p.q. is $-x/10$, and the convergence ratio is thus 0.4. This ratio is clearly seen in the last two p.r.'s above; and we take opportunity to remark, without going into details of proof, that the coefficients in the p.r. may be shown to have properties, in regard to convergence and the vector of errors, quite similar to those of the r.p.r.

The new feature is that the errors in the coefficients of the last r.p.r., namely -0.00265 , -0.00530 , -0.00795 , are visibly in ratio 1:2:3. Or we may take the *differences* of corresponding coefficients in the last two r.p.r.'s, 0.00398 , 0.00793 , 0.01190 , which show the same approximate ratio. The reason is (§ 7) that $x^3 + 2x + 3$ is an exact factor of the divisor $x^3 + 6x^2 + 11x + 12$; the other factor $x + 4$ has been used up, as it were, in producing the convergence.

6. PROCESSES FOR THE ACCELERATION OF CONVERGENCE

It has already been remarked that the nature and rapidity of convergence depend on the latent root ρ_1 of R , of greatest modulus. If ρ_1 is real, and sufficiently separated in modulus from ρ_2 , then the first differences of corresponding coefficients in the sequence of r.p.r.'s settle down to an approximate geometrical progression. The computer, on seeing indications of this in the work, will then find it advantageous to stop, to form first and second differences of corresponding coefficients in at least three consecutive r.p.r.'s, and to use these in the following way to obtain superior approximation.

Let three consecutive members of such a sequence of coefficients be named u_{k-1} , u_k , u_{k+1} . Then a better sequence (Aitken, 1926) is given by

$$u_k^{(1)} = - \left| \begin{array}{cc} u_k & u_{k+1} \\ u_{k-1} & u_k \end{array} \right| \div (u_{k+1} - 2u_k + u_{k-1}), \quad (1)$$

which is very easily computed if a machine is available. A useful variant (Steffensen, 1933) is to construct

$$u_k^{(1)} = u_{k+1} - (\Delta u_k)^2 / \Delta^2 u_{k-1}. \quad (2)$$

Yet a further equivalent, which can serve as a check upon calculation (not upon convergence), is

$$u_k^{(1)} = u_k - \Delta u_k \Delta u_{k-1} / \Delta^2 u_{k-1}. \quad (3)$$

A check on *convergence* is given either by using the similar improvement on a neighbouring triad such as u_{k-2} , u_{k-1} , u_k , or by trying out the process on the whole r.p.r. and taking the resulting polynomial as a new divisor.

Example 1.—If the working in Wallis's equation, Example 1 of § 5, had been taken to ten significant digits, the last three approximations to the real root s , with first and second differences, would have been:

	Δ	Δ^2	check by (3)
0.09454979283			
	178215		
9455157498		- 188076	9455157498
- 9861			178215 \times 9861
9455147637			- 9344 = $\frac{\quad}{188076}$
			9455148154
517 = 9861 ² / 188076			
<hr/> 9455148154			

The improved value, obtained both by Steffensen's adjustment (2) and by the variant (3), is shown. Here convergence is so good that it is hardly necessary to use an accelerative process. The true value (Whittaker and Robinson, 1929, p. 106, footnote) is 0.094551481542 . . .

Example 2.—We take the values of the coefficients in the last four r.p.r.'s of Example 2, § 5, and apply the adjustment (2):

Δ	Δ^2	Δ	Δ^2	Δ	Δ^2
6.92150		13.77132		7.84982	
4667		13454		8787	
6.96817	-2767	13.90586	-7861	7.93769	-5094
1900		5593		3693	
6.98717	-1142	13.96179	-3312	7.97462	-2180
768		2281		1513	
6.99485		13.98460		7.98975	
516 = 768 ¹ / 1142		1571 = 2281 ² / 3312		1050 = 1513 ³ / 2180	
7.00001		14.00031		8.00025	

The improved values 7.0000, 14.0003, 8.0003 are close to the true integer values.

Oscillatory Convergence.—If convergence is dominated by two conjugate complex roots of R , corresponding to a certain pair of conjugate complex roots of the exact factor that is being found (and we recall that this need not be the factor of $f_n(x)$ having the numerically smallest roots), then we may apply (Aitken, 1926) an accelerative process slightly more complicated than (1) but equally effective. We take the opportunity here of transforming it into an adjustment resembling Steffensen's, as follows.

Consider a sequence . . . , u_{k-1} , u_k , u_{k+1} , . . . the first differences of which are settling down, because of the presence of a dominant pair of complex roots in R , to a sequence like the following:

$$\dots, a \cos \alpha, ar \cos(\theta + \alpha), ar^2 \cos(2\theta + \alpha), ar^3 \cos(3\theta + \alpha), \dots \quad (4)$$

where r (< 1) is the modulus and θ the amplitude of the roots in question, and α is a constant angle induced by the initial conditions. Then it is not difficult to show that the sequence of persymmetric determinants of the second order, easily evaluated by machine,

$$\dots, \begin{vmatrix} \Delta u_{k-1} & \Delta u_k \\ \Delta u_{k-2} & \Delta u_{k-1} \end{vmatrix}, \begin{vmatrix} \Delta u_k & \Delta u_{k+1} \\ \Delta u_{k-1} & \Delta u_k \end{vmatrix}, \dots \quad (5)$$

is tending to a geometric progression of common ratio r^2 , whence a sufficient approximation to r^2 may be found for practical use. The desired adjustment is then

$$u_k^{(1)} = u_{k+1} - \frac{\Delta u_k (\Delta u_{k+1} - r^2 \Delta u_k)}{\Delta^2 u_k - r^2 \Delta^2 u_{k-1}}. \quad (6)$$

A variant, analogous to (3) and serving in the same way as an arithmetical check, is

$$u_k^{(1)} = u_k - \frac{\Delta u_k (\Delta u_k - r^2 \Delta u_{k-1})}{\Delta^2 u_k - r^2 \Delta^2 u_{k-1}}. \quad (7)$$

Example 3.—We take the coefficients in the last five r.p.r.'s of Example 4, § 5, and apply the adjustments (6) and (7) to them:

	Δ	Δ^2		Δ	Δ^2
2.01260			3.12528		
	-4185			-10441	
1.97075		4525	3.02087		5892
	340			-4549	
1.97415		1245	2.97538		4516
	1585			-33	
1.99000		-512	2.97505		1444
	1073			1411	
2.00073			2.98916		

The trend of persymmetric determinants made from first differences suggests that r^2 is approximately 0.318. We may remark that the actual value, obtained by substituting the roots of $x^2 + 2x + 3 = 0$ in the r.p.q. $-(x^2 + 14x)/40$, is $513/1600 = 0.3206$.

Substituting the approximation 0.318 in (6) and (7), we have the improved values of the coefficients:

By (6),

$$1.99000 - \frac{0.01585(1073 - 0.318 \times 1585)}{-512 - 0.318 \times 1245} = 1.99993.$$

By (7),

$$1.97415 - \frac{0.01585(1585 - 0.318 \times 340)}{-512 - 0.318 \times 1245} = 1.99994.$$

By (6),

$$2.97538 - \frac{-0.04549(-33 + 0.318 \times 4549)}{4516 - 0.318 \times 5892} = 2.99952.$$

In the case of the second coefficient we cannot make full use of (6) or (7), because of the accidental circumstance, purposely exemplified here, that the second difference 1444 is very close to 0.318×4516 .

7. THE ASYMPTOTIC VECTOR OF ERRORS

We consider next what information can be derived from the vector of errors or, equivalently, of first differences of corresponding coefficients in consecutive r.p.r.'s, at a sufficiently advanced stage. We confine our attention to the case in which the dominant latent root ρ_1 of R is real. It is well known, in problems concerned with determining latent roots and latent vectors of a matrix, that if v is an arbitrary vector then in general

$R^t v$ tends with increasing t to the dominant latent vector, so that $R^{t+1} v$ tends to $\rho_1 R^t v$. Hence in the present application the ratios of the first differences of coefficients from one r.p.r. to the next will tend to the latent vector shown in § 4 (1). But the elements of this vector are coefficients of powers of x in a certain divisor of $f_n(x)$ and of $d_m(x)$, namely the polynomial of degree $m-1$:

$$d_{m-1}(x) = d_m(x)/(x - \beta_1), \quad (1)$$

where β_1 corresponds to ρ_1 .

Thus not only does iterated penultimate remaindering disclose the factor $d_m(x)$, but by inspection of vectors of differences in consecutive r.p.r.'s we discern the further factorization of $d_m(x)$ into $(x - \beta_1)d_{m-1}(x)$. This is so useful an addition to the existing resources that it is well worth while, if a machine is at hand, to preserve two or three more digits in the entries, so that the differences can be more accurately determined. Under suitable conditions the limiting ratios of these differences can be found still more exactly by the use of accelerative processes like those described in § 6.

Example 1.—The ratios of corresponding first differences in Example 2 of § 5 are shown below. They are not very well determined, since there are only three or four digits in the later differences; we shall, however, apply the accelerative process to them, obtaining a fair approximation to the coefficients of the true divisor $x^3 + 3x + 2$.

Differences :	4667	13454	8787	Ratios :	1	2.883	1.883
	1900	5593	3693		1	2.944	1.944
	768	2281	1513		1	2.970	1.970
				Adjusted :	1	2.989	1.989

The first differences of the ratios are themselves seen to be equal. This points to the further factor $x + 1$.

Example 2.—The above result continues to hold when the zeros of $d_{m-1}(x)$ contain conjugate complex pairs among them, always provided that the dominant latent root of R is real. Below we show the ratios of first differences taken from Example 4 of § 5. If more digits had been kept in the working, the accelerative process of § 6 (6) could have been applied.

Differences :	2507	5046	7649	Ratios :	1	2.013	3.051
	994	1981	2991		1	1.993	3.009
	398	793	1190		1	1.992	2.990

An oscillating tendency towards the coefficients of the exact divisor $x^3 + 2x + 3$ is clearly seen.

8. GENERAL OBSERVATIONS AND CONCLUSIONS

The theory of the iterated penultimate remainder, as expounded in the preceding sections, has reference to those stages of the process at which the errors ϵ_i have become relatively small enough for their squares, cubes or products of higher than the first degree to be neglected. But at the outset, in choosing a first trial divisor, we cannot know how small the errors are; they may indeed, and often will be, much too large for their higher powers to be neglected. The discussion of the r.p.r. sequence in such an unrestricted case would require a more general and a more difficult analysis; and it would almost certainly be hedged about by a multitude of special conditions and exceptions. We have ourselves examined it only slightly, but we dare assert in a general way, after the study of numerous arithmetical examples, that even in cases where the initial divisor is almost fantastically remote from the final exact divisor an initial apparent tendency to diverge may be changed in a few steps to a surprising convergence. Any reader sufficiently interested may test this assertion by trying, on such a polynomial as $x^3 + 13x^2 + 34x + 40$, so wild a divisor as $x^3 + 100x + 500$, or the like.

It is possible that the situation could be expressed in geometrical language somewhat as follows. The set of coefficients in a trial divisor may be regarded as a vector or point in space of m dimensions. In this r.p.r. space the vectors corresponding to exact factors are points of equilibrium, stable or unstable as the case may be, with respect to the operation of penultimate remaindering. Considering for example the case of quadratic divisors, intuition suggests that the regions of stability in the r.p.r. plane are separated by critical curves, along which there is, so to speak, equal attraction towards one or other stable point. The strength of stability at a point could be measured by the reciprocal of the associated $|\rho_1|$ for that region. As we have mentioned more than once, it is not to be assumed, even for linear divisors, that the divisor corresponding to the zero of $f_n(x)$ of smallest modulus corresponds to the point of greatest stability. This can be seen from such an example as

$$f_3(x) = (2x - 1)(4x - 1)(8x + 1). \quad (1)$$

The divisors $4x - 1$, $8x + 1$ will be found to be stable, while $2x - 1$ is unstable. However, though $4x - 1$ corresponds to a root of larger modulus than $8x + 1$ does, yet $4x - 1$ is the more stable; it will in fact be found from § 4 (5) that the convergence-ratios of the iteration are $-\frac{1}{2}$ for the divisor $4x - 1$ and $-\frac{7}{8}$ for $8x + 1$. It is not difficult, by choosing roots with suitable variations of sign, to construct polynomials $f_n(x)$ having different

degrees of stability among the quadratic divisors, the divisor corresponding to the roots of least modulus being less stable than other divisors. We may also suggest that besides critical curves separating the regions of stability there will be singular points. For example, $x^3 + 8x^2 + 16x + 5$ has the unstable divisor $x + 5$; yet the trial divisor $x + 3$ gives the p.r. $x + 5$, exact. The reason is that the penultimate *quotient* contains both $x + 5$ and $x + 3$ as factors.

It has appeared, therefore, that the method of iterated p.r. will not necessarily, nor usually, give the roots of a proposed algebraic equation in either ascending or descending order of moduli. This is one of the features that distinguish it from extant methods such as the root-squaring, or the extended Bernoulli method.

The next question that arises is how to proceed if the r.p.r.'s show evident signs of divergence. This will again depend, under the assumptions we made in § 2, on the nature of the latent roots of R . It may be (though naturally this could hardly ever be known beforehand) that there are latent roots of R with modulus > 1 . In such a case we might try the reciprocal equation; but even this might not be successful. The remedy might then be to give up trying for a divisor of degree m , and to try others of degree $m - 1$ or $m + 1$. Indeed a pedestrian but often effective procedure is to try first of all for linear or quadratic factors, and to divide them out from $f_n(x)$ as soon as they are sufficiently determined. However, most of these points are matters requiring practical experience. In any case the method of iterated p.r. is a useful addition to the existing methods of solving algebraic equations, and would probably lend itself very well to "programming" on high-speed electronic machines.

In conclusion, we hazard a few remarks regarding Friedman's method. This, though doubtless suggested in idea by Lin's examples, must be regarded as essentially distinct and new in principle. It is also, one can hardly doubt, more complicated in theory. Friedman's procedure is as follows: to choose a trial divisor and to divide $f_n(x)$ by it, right to the final remainder stage; then, discarding the remainder, to take the quotient and to divide $f_n(x)$ again by this, but according to *ascending* powers of x , until a second quotient, of the same degree as the first trial divisor, is reached. This second quotient, under certain conditions, is closer to an exact factor of $f_n(x)$ than the first trial divisor was. Friedman's proposal is to iterate this two-way division and to make a sequence of the second, the backward quotients. Now the coefficients in the forward quotient are (§ 2) bigradients; those in the backward quotient will also be bigradients, but *compound* ones, their elements being the bigradients of the earlier set; and in the final reduced quotient a division by the leading coefficient, itself a

compound bigradient, will have to be applied. It is possible, and examination confirms it, that the determinant quotients may be simplified to a certain extent; yet even so the situation remains complicated. Friedman, confining his discussion to linear and quadratic divisors, concludes that convergence in these cases, when it exists, is superior to that of Lin's method. This may be so in some cases, but it is clear that the number of operations required to obtain each iterate is a shade more than twice that required in penultimate remaindering, and the change-over to a reversed division makes a break of rhythm in the running routine. It is certain also—this can be verified by trying Friedman's method on equation (1) above—that each method can converge in circumstances where the other diverges.

REFERENCES TO LITERATURE

- AITKEN, A. C., 1926. *Proc. Roy. Soc. Edin.*, XLVI, 289-305.
 —, 1931. *Proc. Roy. Soc. Edin.*, LI, 80-90.
 —, 1937. *Proc. Roy. Soc. Edin.*, LVII, 269-304.
 FRIEDMAN, B., 1949. *Comm. Pure and Appl. Math.*, II, 195-208.
 FRY, T. C., 1945. *Quart. Appl. Math.*, III, 89.
 LIN, S. N., 1941. *Journ. Appl. Math. and Phys.*, XX, 231-241. (The method outlined here seems to have been first used in a thesis submitted for the degree of Sc.D. to the M.I.T. in 1939.)
 STEFFENSEN, J. F., 1933. *Skand. Akt.-Tidskr.*, 64-72.
 WHITTAKER, E. T., and ROBINSON, G., 1929. *The Calculus of Observations*, 2nd Ed., 79-83, 106.

XIV.—Experiments in Diffraction Microscopy.* By G. L. Rogers,†
M.A., Ph.D., Department of Physics, University College,
Dundee, Angus. *Communicated by Professor G. D. PRESTON.*
(With Two Plates and Nine Text-figures)

(MS. received October 31, 1950. Read May 7, 1951)

SYNOPSIS

Experiments have been performed, using purely optical methods, to verify and extend the theory of Gabor's diffraction microscope. An elementary theory of the process is first given, from which certain generalizations are provisionally drawn. In particular, a focal length is attributed to any Fresnel diffraction pattern and the hologram derived from it by photography. The variation of this focal length with wavelength and scale factor is postulated by analogy with a zone-plate, and the power-rate for a hologram is defined. These deductions are then verified by experiment, and a summary is given at the end of §10. Various other confirmatory experiments are then described.

Adequate information is given about apparatus and technique to enable new entrants into this field to obtain satisfactory results with the minimum of preliminary trial.

§ 1. INTRODUCTORY

In a number of papers (Gabor, 1948, 1949) Gabor has described a system of microscopy in coherent light whereby a magnified but indistinct image of an object may be obtained without a lens, and a "reconstructed" or more distinct image later produced with an auxiliary lens. In particular, it is proposed to perform the first stage electronically, with a lensless electron microscope, and the second stage optically with a luminous source and suitable reconstructing lens.

The present paper describes an experimental study, using optical methods throughout, of this very ingenious idea. Though the application to electron microscopy has not been entirely forgotten, the work has been directed to the method in its own right. As a result a number of useful generalizations have been discovered empirically and verified theoretically which, though doubtless implicit in the many equations of Gabor (1949), can also with profit be stated explicitly in simpler physical terms.

§ 2. ELEMENTARY THEORY : FRESNEL DIFFRACTION

(a) CIRCULAR SYMMETRY

Consider a point source of monochromatic light at O and a small scattering object at R . In practice, a dust particle or fine droplet serves

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

† Since appointed Lecturer at Victoria University College, Wellington, New Zealand.

admirably. Let us consider the Fresnel diffraction pattern at a plane PM due to the direct radiation from O and scattered radiation from R (fig. 1). We take OR perpendicular to PM , with M on OR . Also let $OR = a$, $RM = b$, in accordance with the notation generally used in Fresnel diffraction.

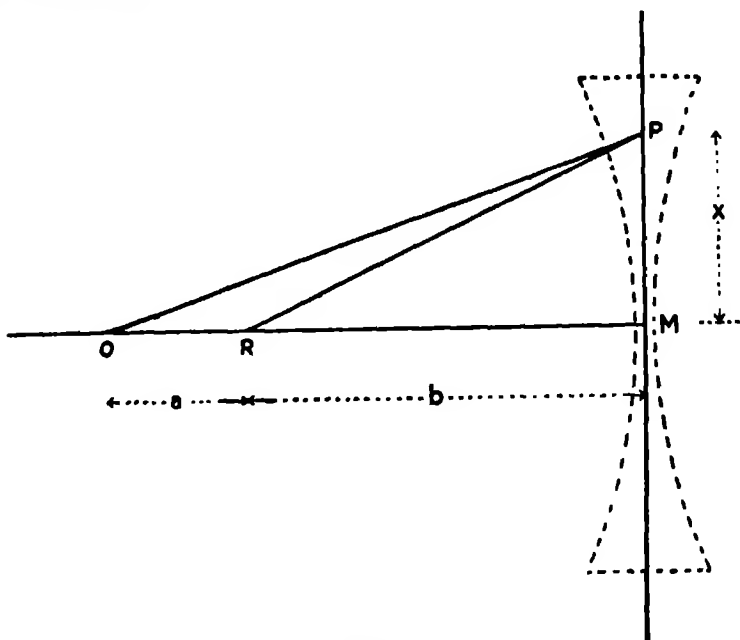


FIG. 1

The direct ray travels along a path OP . If we take $PM = x$, we get :

$$OP^2 = (a + b)^2 + x^2 = OM^2 \left(1 + \frac{x^2}{(a + b)^2} \right).$$

Suppose now $x \ll OM$, then :

$$\begin{aligned} OP &= OM \left(1 + \frac{x^2}{2 \cdot OM^2} + \text{higher terms in } \frac{x^2}{OM^2} \right) \\ &\simeq OM + \frac{x^2}{2 \cdot OM} = a + b + \frac{x^2}{2(a + b)}. \end{aligned}$$

Similarly

$$RP \simeq RM + \frac{x^2}{2 \cdot RM} = b + \frac{x^2}{2 \cdot b}.$$

If now we allow for the fact that the scattered light has to go from O to R before it is scattered, and if we assume zero phase-change on scattering, we get the path from O to P via R , viz.,

$$OR + RP = a + b + \frac{x^2}{2b}.$$

The path difference between the two disturbances will then be :

$$\begin{aligned}\Delta &= a + b + \frac{x^2}{2b} - a - b - \frac{x^2}{2(a+b)} \\ &= \frac{x^2}{2} \left[\frac{1}{b} - \frac{1}{a+b} \right].\end{aligned}$$

It is now convenient to replace the bracket by the parameter $1/f$ according to the equation :

$$\frac{1}{b} - \frac{1}{a+b} = \frac{1}{f}.$$

It will be seen at once that this parameter has a physical meaning, which experiment shows to be fundamental to the whole process. In short, f is numerically equal to the focal length of that divergent lens which, placed in the screen position, would image the source in the plane of the object.

We get now :

$$\Delta = \frac{x^2}{2} \cdot \frac{1}{f}.$$

Since the intensity of illumination at P is a maximum when Δ is an even number of half wavelengths, and minimum if Δ is an odd number of half wavelengths, we get that :

$$\frac{x^2}{2} \cdot \frac{1}{f} = p \cdot \frac{\lambda}{2} \quad p \text{ even} \rightarrow \text{max.}$$

$$\frac{x^2}{2} \cdot \frac{1}{f} = p \cdot \frac{\lambda}{2} \quad p \text{ odd} \rightarrow \text{min.}$$

This gives $x^2 = fp\lambda$ as defining a series of rings about M , with maxima at even values of p and minima at odd values of p . The regions between will have intermediate intensities varying in an approximately sinusoidal manner.

If the scattering particle, R , introduces a phase shift, the result is an opening out or closing in of the system of rings, corresponding to the addition of a constant term to Δ . If, for example, there is a phase delay of $\frac{1}{2}\pi$, corresponding to a path increase of $\frac{1}{2}\lambda$, we get :

$$\Delta = \frac{x^2}{2} \cdot \frac{1}{f} + \frac{\lambda}{4} = p \frac{\lambda}{2}$$

and

$$\frac{x^2}{2} \cdot \frac{1}{f} = p \cdot \frac{\lambda}{2} - \frac{\lambda}{4} = q \cdot \frac{\lambda}{2}$$

where

$$q = p - \frac{1}{2}$$

The first minimum occurs at $p = 1$ or $q = \frac{1}{2}$ and the first maximum occurs at $p = 2$ or $q = 1\frac{1}{2}$. Thus the rings defined by $q = 1, 2, 3$, etc., which in the previous case were minima or maxima, now become rings of average intensity. Suppose now we replace the sinusoidal variations of intensity by abrupt changes of intensity at the positions $q = 1, 2, 3$, etc., as the intensity moves through the average, so that $I = 0$ from $q = 0 \rightarrow 1, 2 \rightarrow 3, 4 \rightarrow 5$, etc., and $I = I_{\max}$ when q goes from $1 \rightarrow 2$ or $3 \rightarrow 4$ or $5 \rightarrow 6$.

Such an arrangement, translated from intensities into densities, is the well-known zone plate.

A zone plate is thus a black and white approximation to the Fresnel diffraction pattern of a small object scattering the incident light with a $\frac{1}{2}\pi$ change of phase.

In particular, if x_1 is the radius of the first ring (where $q = 1$) we have:

$$x_1^2 = f\lambda, \text{ or } f = \frac{x_1^2}{\lambda},$$

the well-known expression for the focal length of a zone plate.

Thus it will be seen that f is the focal length of the zone plate, as well as the parameter of the diffraction pattern. We note further that if O is a source of monochromatic light of wavelength λ , and R be removed, the placing of a zone plate of focal length f in the position of the screen will result in the production of an image at the point R of the source O . The zone plate may thus be said to "reconstruct" the object R in the correct position.

We shall see in general that if the Fresnel diffraction pattern of an object is photographed or otherwise reproduced, and then placed in the original diverging beam, it "reconstructs" the object from the point source in its original position. (This constitutes the Gabor method of diffraction microscopy.) But further, if the reproduction of the Fresnel diffraction pattern (called by Gabor the "hologram") is placed elsewhere in a diverging beam, it will produce a reconstruction of the object in a position related to the source position by the lens formula.

Two other properties of the zone plate deserve attention. First, as is well known, the zone plate can act as both a positive and a negative lens. That is to say, its focal length is strictly speaking $\pm f$. Thus not only does it produce an image of O at R , but also an image of O where a convergent lens of focal length f would form an image of O . There are thus *two* reconstructions of the point R . This has been found by Gabor.

Gabor has shown that the process of photographing the diffraction pattern to form the hologram results in a loss of information; information as to the *phase* of the disturbance in the plane PM . (See also

Bragg, 1950.) The assumption in his method is that if the direct beam is of much greater intensity than the scattered beam, the phases of the resultant disturbance will nowhere differ greatly from that of the direct beam. As the latter is the phase supplied in the reconstruction, this loss of information is, to the first approximation, unimportant. He goes on to show that, as a second approximation, this loss of information results in the production of a secondary image on the side of O remote from R : making the approximation that $OR \ll RM$, he gets the secondary image at R' where $R'O = OR$. This also holds for a zone plate of focal length $\gg OM$. It is clear that the double sign of the zone plate focal length is equivalent to the phenomenon of the hologram in producing two images.

The second property of the zone plate worthy of study is its possession of a whole series of secondary focal lengths, corresponding to higher



Fig. 2.—Circular zone plate with third order missing from its higher order foci

orders in a diffraction grating. The *powers* of these lenses go up as the odd integers, 1, 3, 5, etc., and may be regarded as higher orders of the fundamental power. It will be seen that the even orders are missing, as in the case of a grating with equal black and white stripes. The reason is in both cases the same. In the first order position, the transparent areas in both devices transmit a range of phases from 0 to π , 2π to 3π , etc., which reinforce. But in all even orders, the range of phases transmitted runs from 0 to $2n\pi$, $4n\pi$ to $6n\pi$, etc., and darkness results. If a zone plate were constructed to transmit only the first third-period zone, blocking the next two-thirds and so forth, it would be found to possess all powers bar multiples of 3. We have verified that all multiples of 3 are absent from the zone plate in fig. 2.

The reason for this lies in the fact that the x values correspond to integral values of p in $\sqrt{(pf/\lambda)}$ and hence, say, a trebling of p throughout gives a similar set of x values corresponding to a new focal length

$f' = f/3$. This property may be peculiar to systems of circular or linear symmetry where the x 's recur at regular values of \sqrt{p} .

§ 3. ELEMENTARY THEORY : FRESNEL DIFFRACTION, (b) STRAIGHT LINE SYMMETRY

In the case of an object with a straight line symmetry, a so-called "cylindrical" wavefront is often taken, though to the degree of approximation normally used in these discussions similar results occur with spherical wave-fronts impinging on the straight line object. Thus Jenkins and White (1937), using cylindrical wavefronts and Airy (1877), using spherical wavefronts, both arrive at Fresnel Integrals or the equivalent Cornu spiral as the correct method of solving these problems.

The Fresnel Integrals are tabulated in terms of a dimensionless parameter, v (say), and the minima and maxima of a particular problem may be obtained from the tables in terms of v . In the case of Cornu's Spiral, however, owing to the fact that we plot dimensionless numbers on some particular linear scale, the parameter v acquires the appearance of linear dimensions, and is said to be the length along the arc of the spiral. But this arises from the fact that the scale of the axis is itself a dimensional quantity, to which v is strictly proportional; a fact often overlooked in the discussion of graphical problems.

If O (fig. 1) is the source and R the linear object producing a Fresnel diffraction pattern in the plane PM , we get the positions of minima and maxima in terms of the v values in the tables. These have to be translated into x values in the PM plane. If $OR = a$, $RM = b$, it can be shown that for a wavelength λ , the relation between x and v is (Jenkins and White):

$$x = v \sqrt{\left(\frac{b\lambda(a+b)}{2a}\right)}.$$

If, now, we attempt to express this relation in terms of our parameter, f , we see at once:

$$\begin{aligned} \frac{1}{f} &= \frac{1}{b} - \frac{1}{a+b} = \frac{a+b-b}{b(a+b)} = \frac{a}{b(a+b)} \\ f &= \frac{b(a+b)}{a} \end{aligned}$$

and in particular:

$$x = v \sqrt{\left(\frac{\lambda}{2} \cdot \frac{b(a+b)}{a}\right)} = v \sqrt{\left(\frac{\lambda f}{2}\right)}$$

We see, therefore, that the *scale* of the pattern is, as before, a linear function of $\sqrt{(\lambda f)}$ where f is the focal length of a lens, placed at the screen, to image the source on the object. The scale thus depends on the value of f rather than on the individual values of a and b .

§ 4. THE LINEAR ZONE PLATE

It is possible to divide up a cylindrical wave-front into a set of linear half-period zones, and thus get a linear zone plate analogous to a cylindrical lens.

There are, in practice, three ways of doing this wave-front division. In the first, the wavefront is left open from the centre of the wave to

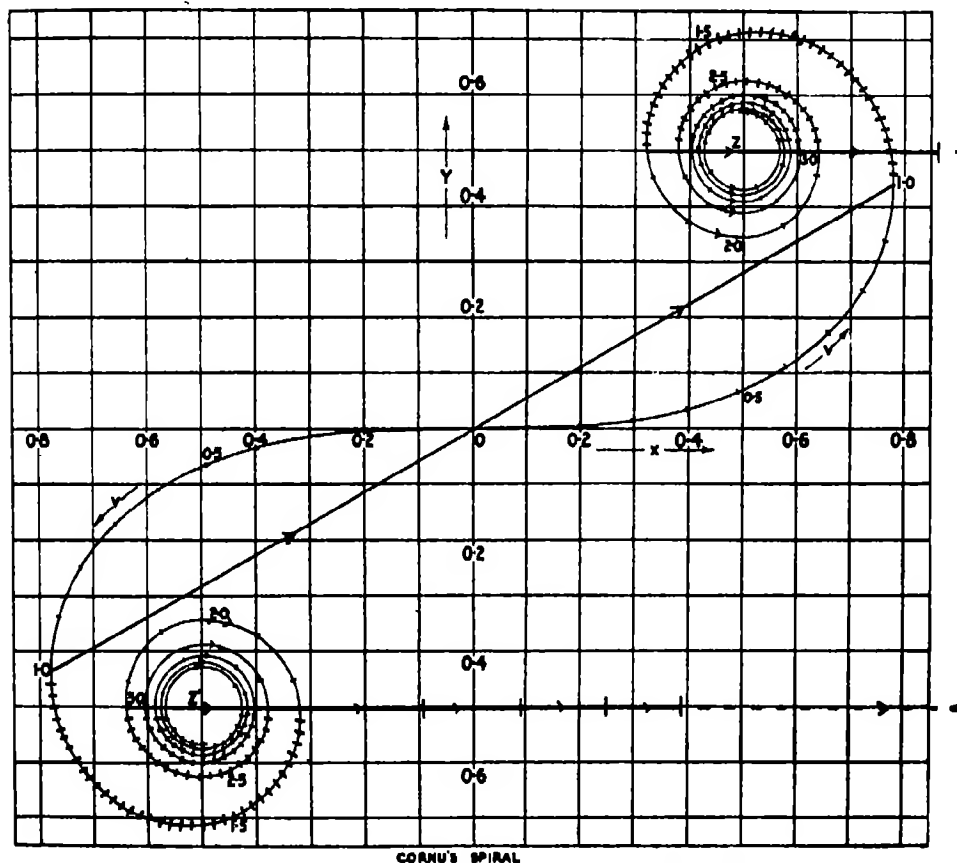


FIG. 3.—Cornu's spiral for type I linear zone plate

points where the disturbance is $\frac{1}{2}\pi$ out of phase (fig. 3). This occurs when the tangent to the Cornu spiral first becomes vertical, and this point occurs when $v=1$. The next zone, which is blackened, runs from here to the next position with a vertical tangent, at a phase angle of $\frac{3}{2}\pi$ and a v value of $\sqrt{3}$. As the angle ψ of the tangent is always given by $\psi = \frac{1}{2}\pi v^2$ we get zone boundaries at v values given by:

$$v^2 = 1, 3, 5, 7, 9, \text{ etc.}$$

With this system of zone boundaries, we get, for one-half of the zone plate, a vector from the origin of the spiral to the point where $v = 1$, which lies at an angle $\tan^{-1} \frac{0.4383}{0.7799}$ to the x axis, and then a set of substantially horizontal vectors arising from the outer zones. The lengths of these vectors can be obtained approximately from the radius of curvature formula of the spiral:

$$\rho = \frac{1}{\pi v}.$$

For example, the open zone from $v^2 = 3$ to $v^2 = 5$ can be regarded as having an average v^2 value of 4, giving $v = 2$ and $\rho = 1/\pi\sqrt{4}$. The length of the $3 \rightarrow 5$ vector will thus be $2\rho = 2/\pi\sqrt{4}$. Similarly the length of the $7 \rightarrow 9$ vector will be $2\rho = 2/\pi\sqrt{8}$. Thus the sum of the outer zones:

$$S_{\text{outer}} = \frac{2}{\pi} \left[\frac{1}{\sqrt{4}} + \frac{1}{\sqrt{8}} + \frac{1}{\sqrt{12}} + \dots + \frac{1}{\sqrt{(4m)}} + \dots \right]$$

or

$$S_{\text{outer}} = \frac{1}{\pi} \left[\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \dots \right].$$

But the above sum is divergent. Thus, in the absence of an obliquity factor, these horizontal vectors add up to an indefinitely large amount, and ultimately swamp the vector from the inner zone. The system thus gives a maximum on the axis at the chosen point.

When we examine the possible higher orders of this system, we find that the even orders vanish, because the outer zones contribute regions which are substantially closed circles on the Cornu spiral, leaving the initial vector of length similar to that of the unobstructed path. Odd orders will be present, as the open areas then correspond to an odd number of $\frac{1}{4}$ turns in the spiral giving a divergent series of substantially horizontal vectors.

The second type of zone plate obtainable is that which results by considering the intersection of the Cornu spiral with the line $x = y$. This was the system originally chosen for practical experiments, as all the vectors lie along $x = y$ in the focal position. Also, it gives a maximum value for the initial vector. The diagram for this type is shown in fig. 4. A finite zone-plate of this type is shown in fig. 5.

In order to get the " v " values we have to solve the equation $C(v) = S(v)$ where $C(v)$ and $S(v)$ are the Fresnel integrals. This can be done by inspection of the tables, but approximate roots can be readily obtained. For this, we assume that Cornu's spiral always crosses the line $x = y$ at right angles. Although not strictly accurate for the first turn or two, the approximation becomes rapidly better as v increases.

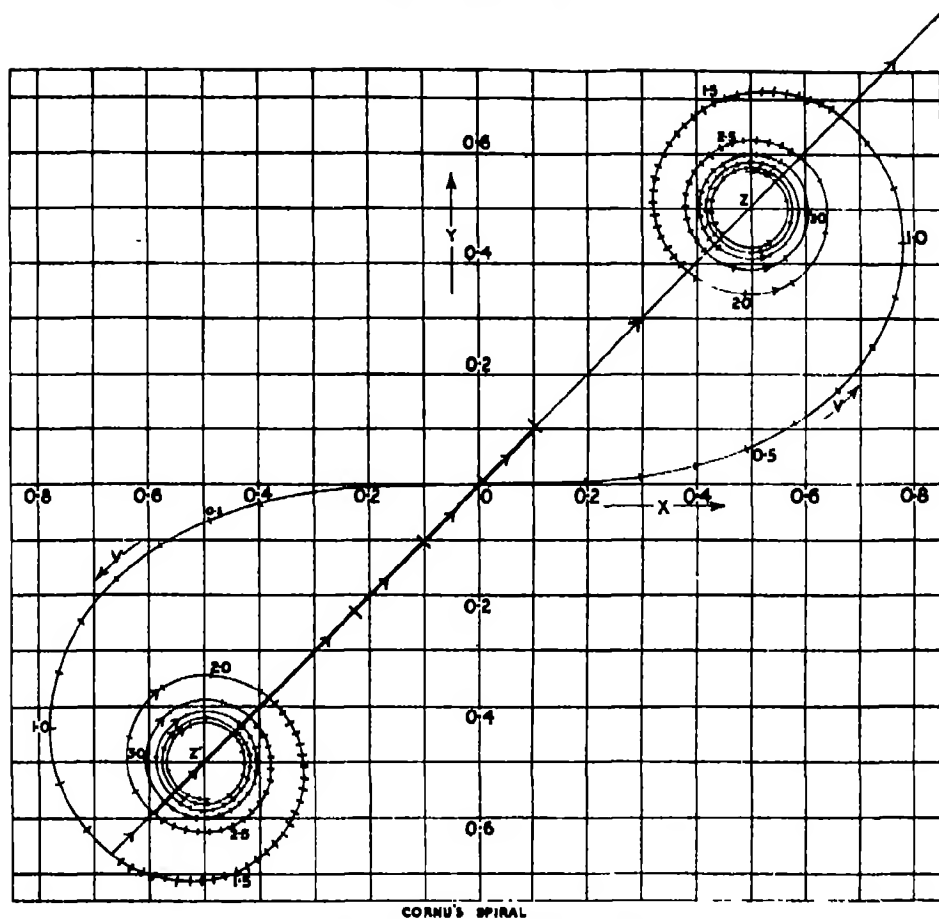


FIG. 4.—Cornu's spiral for type II linear zone plate

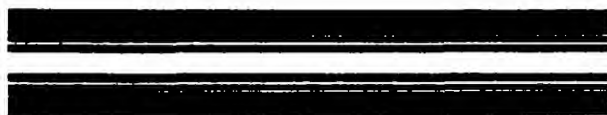


FIG. 5.—Type II linear zone plate

Since the slope of $x = y$ is $+1$ this means we have to find the points on Cornu's spiral where the slope is -1 .

If $\tan \psi = -1$, then $\psi = \frac{3}{4}\pi \pm n\pi$ and we also have $\psi = \frac{1}{4}\pi v^2$. Hence the solution of the equation is:

$$v^2 = 2n - 1$$

OR

$$v^2 = 1\frac{1}{2}, 3\frac{1}{2}, 5\frac{1}{2}, \dots, \text{etc.}$$

We now have the range 0 to $v^2 = 1\frac{1}{2}$ open, close $v^2 = 1\frac{1}{2}$ to $v^2 = 3\frac{1}{2}$ and open $v^2 = 3\frac{1}{2}$ to $v^2 = 5\frac{1}{2}$.

This latter gives a vector along $x = y$ to add to the first vector. Taking $4\frac{1}{2}$ as the average value of v^2 , we get its length to be $2\rho = 2\sqrt{2}/\pi\sqrt{9}$.

The outer vectors then become :

$$\frac{2\sqrt{2}}{\pi} \left[\frac{1}{\sqrt{9}} + \frac{1}{\sqrt{17}} + \frac{1}{\sqrt{25}} + \dots + \frac{1}{\sqrt{(8m+1)}} + \dots \right].$$

This, again, is a divergent series indicating a bright line corresponding to the first order.

The high order effects with this grating differ a little from those of the first type. Even orders are missing, as with the other types and for the same reason. The outer zones make little or no contribution, being virtually closed circles.

The first order which produces vectors along the line $x = y$ is the fifth order. The third order produces vectors from the outer zones which lie substantially at right angles to the line $x = y$. As this series is also divergent, we must suppose that this vector will swamp the original vector, and give a bright line with a phase $\frac{1}{2}\pi$ different from the first order.

The fifth order series is opposed to the initial vector, and must swamp it, giving a disturbance π out of phase with the first order disturbance. With a finite zone plate, of the right number of terms, the 5th order system may give darkness, the finite series from the outer zones being equal and opposite to the initial vector.

Similarly the 7th order gives a system of vectors perpendicular to $x = y$ and $\frac{3}{2}\pi$ out of phase with the initial vector. The 9th order is the first where the outer vectors are parallel to the initial vector and in the same sense.

We have shown that the sum of the outer vectors for the first order is

$$\frac{2\sqrt{2}}{\pi} \left[\frac{1}{\sqrt{9}} + \frac{1}{\sqrt{17}} + \frac{1}{\sqrt{25}} + \dots + \frac{1}{\sqrt{(8m+1)}} + \dots \right].$$

Similarly, by considering the average radius of the approximately circular arcs in the 5th order, we get a sum :

$$\frac{4}{\pi\sqrt{10}} \left[\frac{1}{\sqrt{9}} + \frac{1}{\sqrt{17}} + \frac{1}{\sqrt{25}} + \dots + \frac{1}{\sqrt{(8m+1)}} + \dots \right].$$

The ratio of these two vectors is $1 : \frac{1}{\sqrt{5}}$. In general the n th order will give rise to an outer vector sum of magnitude $\frac{1}{\sqrt{n}}$ of that of the first order.

We thus see that the *intensities* fall off as the order, and become low at high orders.

The third type of zone plate has an initial zone going a little beyond the maximum initial vector from the origin to the first intersection with $x = y$. Here we allow the 1st or central zone to go out to the point

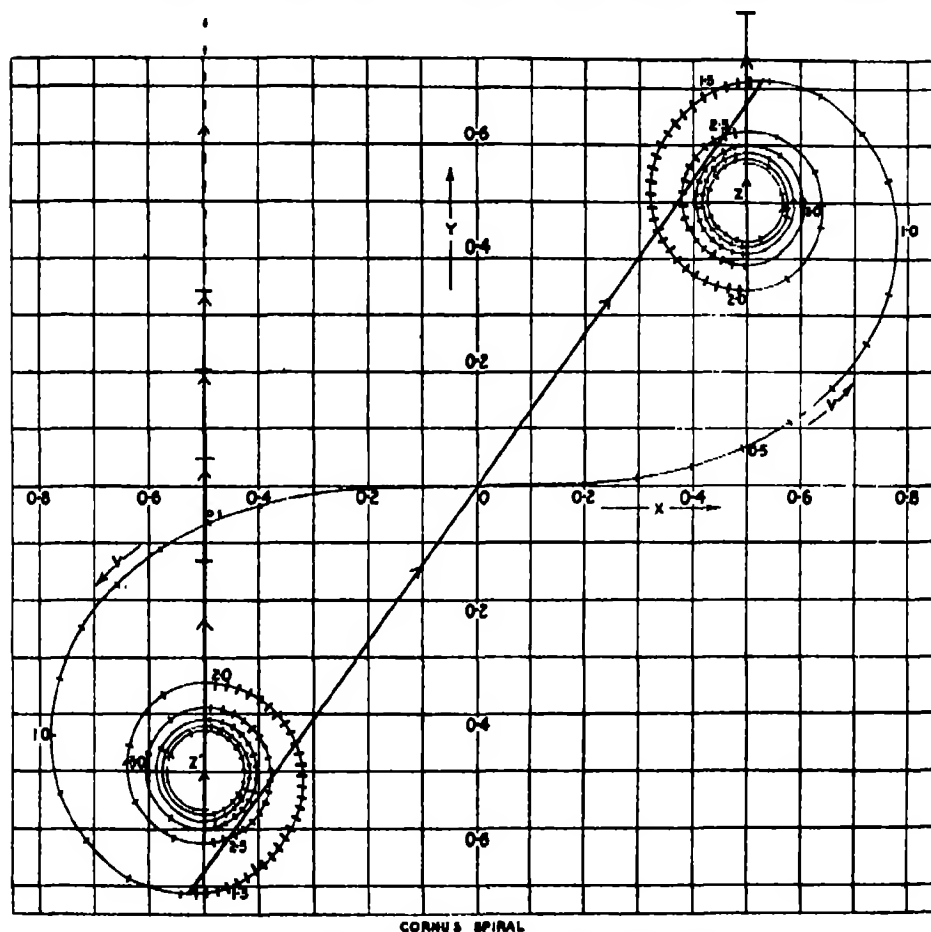


FIG. 6.—Cornu's spiral for type III linear zone plate

where the phase is π , i.e., where the Cornu's spiral is next *horizontal* (fig. 6). This occurs, of course, where $\frac{1}{2}\pi v^2 = \pi$ and gives $v^2 = 2$ as the edge of the zone. Next blot out the zone from $v^2 = 2$ to $v^2 = 4$ and leave open $v^2 = 4$ to $v^2 = 6$. This latter gives a vertical vector directed upwards. We thus get a series, with zone edges at $v^2 = 2, 4, 6, 8 \dots$ etc. Dividing throughout by 2, the ratio of these distances (in v) is as the square roots of the natural numbers, as in the circular zone plates. Once again, the series is divergent, indicating a swamping of the initial vector.

Even orders once again disappear and odd orders make their appearance, but without alternations in phase.

The use of a linear zone plate with a finite number of terms may require special consideration. In fact, experimental work has shown that with a suitable finite number of zones, black lines can be obtained instead of white ones. The detailed consideration of these problems is deferred until a later paper.

§ 5. APPARATUS

In the early work an ordinary Osira lamp was used, as light source, with the glass bulb removed, and the whole in a box to protect the operator from U.V. light. An image of the arc was focused by a 16 mm. microscope objective on to a small hole 10–12 μ diameter of approximately circular outline made by Boy's method (Strong, 1938). This is not an entirely satisfactory procedure, because it is difficult to produce a very fine hole, but it does very greatly reduce the number of spurious fringes otherwise produced by dust in the system. The main difficulty lay, however, in the moderate intrinsic brilliance of the Osira lamp.

Later work was therefore carried out with one of the recently developed high-pressure mercury lamps, running at 250 watts, and developing the greater part of the light in a region between the electrodes of 3 to 4 mm. diameter. The plasma in the Osira filled a tube 1 by 3 cm. This compact source was run in the region of a disk carrying a number of pin-holes of various diameters, from 20—1,000 μ , which disk could be rotated from outside the lamphouse to bring each hole into play in turn as required. The pin-holes allow a cone of light to pass, and this falls on to one side of a microscope objective, either directly or after passing through a colour filter mounted on another disk. The distance of hole to objective is the "working distance" of the objective, and the side of the lamp-house is made to carry the objective. This arrangement forms a reduced image of the hole, outside the lamp-house, in a convenient position for use (fig. 7).

The "optical bench" consists of a wooden shelf with two metre sticks screwed to it to give an edge against which other units could be slid. These consisted of (i) an object holder, (ii) a plate holder, (iii) a hologram holder, (iv) a 7 in. Aero-Ektar lens, (v) an eyepiece, and occasionally (vi) a small lensless "camera". A night microscope (Rogers, 1948) was also available but was not greatly used when the new arc was put into operation (fig. 8).

(i) The object holder simply consisted of a jig to hold 1 to 4 thicknesses of $\frac{1}{4}$ -in. patent plate glass, on which photographic objects were mounted.

(ii) The plate holder was in fact quite a complex device which was designed to allow a printing frame, loaded in the dark-room with a plate,

to be put in an accurately pre-determined position. This consisted of a flat board resting on the shelf and a vertical front with a hole in it. The frame was held against this front with springs. Provision was made for the insertion of a filter or compensating plate before the frame, if neces-

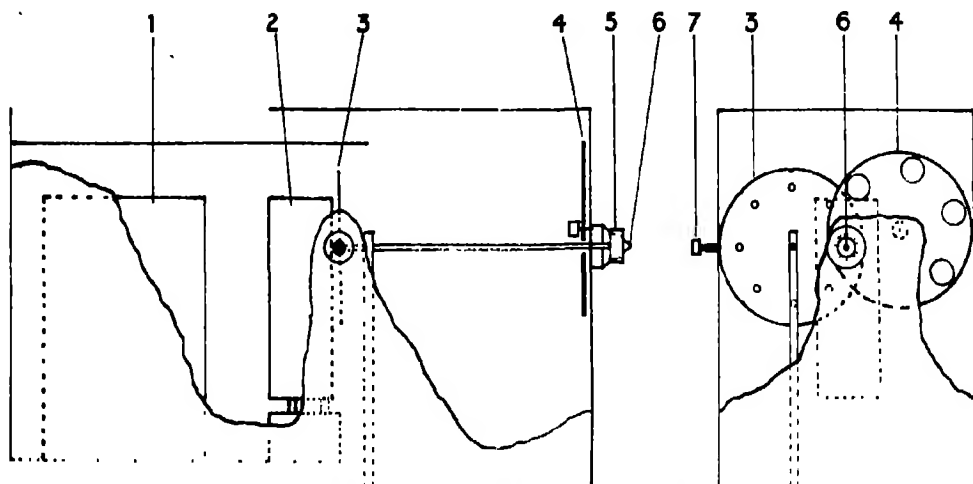


FIG. 7.—Diagram of lamphouse. (1) Choke. (2) Box-type arc. (3) Hole disk. (4) Filter disk. (5) Hole-changing knob. (6) Objective. (7) Device for locating hole disk.

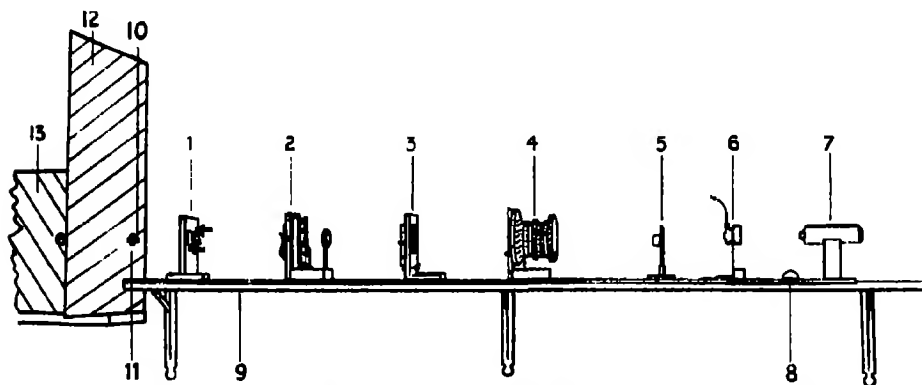


FIG. 8.—Etch Bleach drawing of the apparatus, based on a photograph. (1) Object holder. (2) Plate holder. (3) Hologram holder. (4) Aero-Ektar lens. (5) eyepiece. (6) Lensless camera. (7) Night microscope. (8) Spare objective for night microscope. (9) Shelf. (10) Microscope objective. (11) Knob for changing holes. (12) Screen. (13) Lamphouse.

sary. A medium-focus convex lens, placed behind the frame, was focussed on the plane holding the plate. This thus enabled the whole carriage to be sited in the correct position, the frame being away, by using this lens as a low-powered eyepiece.

(iii) The hologram holder was a simplified version of the above, without lens, designed to hold a hologram against a vertical surface with a hole in it, for convenience in reconstruction. The next two items (iv) and (v) are self-explanatory.

(vi) The small lensless camera consists of a shutter with a circular chamber behind. The back surface of this is a cap, with arrangements for holding a plate the size of a postage stamp. It was originally designed for work with an experimental Bragg "X-ray Microscope", but was useful for photographing the reconstructed images of holograms, when an auxiliary lens was employed.

For normal purposes a run of 2 metres was sufficient, but for certain experiments use was made of a shelf at the same level in an adjacent room, giving in all a possible run of 13 metres.

§ 6. TECHNIQUE

The essence of diffraction microscopy is the production and photography of Fresnel Diffraction patterns. For this purpose monochromatic effects are necessary. The mercury arc gives a number of components in the visible, the most important of which are those at 4358, 5461 and the doublet at 5770—5791 Å. The lines at 4047 and 4078 are, fortunately, comparatively weak, and the two red lines at 6152 and 6232, while useable, are not very strong. There is also a strong U.V. line at 3650 Å. which can fortunately be suppressed easily by any standard U.V. cutting filter.

To obtain a photograph in monochromatic light, it is unnecessary to filter the radiation until it contains only one component. Any number of wavelengths may be allowed to fall on the plate, as long as the plate is sensitive to only one of them. Using a well-known series of filters with sharp short-wave cuts, in combination with a plate of suitable sensitivity, each of these lines can be isolated.

In practice it is found that the following combinations are effective: Ilford "Q" filter to cut the U.V. and partially suppress 4047 and 4078, together with an "ordinary" plate will record only 4358. An ordinary plate contains a pure silver-bromide emulsion without any sensitizing dye. For recording the 5461 it is convenient to use the Wratten 77 or 77A, which not only cuts the blue but suppresses the yellow. If this is used with any orthochromatic plate, the red passed by these filters is quite unimportant. The yellow can be isolated with the Wratten 22 filter. The usual type of orthochromatic emulsion is a little slow to this radiation, but a panchromatic emulsion might record some red. Finally, the red lines can be isolated by an Ilford Narrow Cut Red or a Wratten 26, and recorded on a panchromatic plate. This procedure is slow, and was only used in wavelength dependence tests. Consequently the blue and green lines were most frequently used.

The objects were normally small scale reproductions from black and white line drawings, though one or two attempts to reproduce continuous tone originals have been made. The normal procedure was to remove the shutter from the lensless camera and replace it with a microscope objective screwed into a suitable mount. By successive exposures, it was possible to focus this camera, and when this was done a series of objects would be prepared before disturbing the adjustment. Although technically the objective should have been used at its "working distance" of 160 mm. it was found that quite satisfactory results could be obtained at 1 or even 2 metres. Maximum Resolution plates were used throughout, and the progress of the work was followed under a microscope. The objects, when dry, were cemented with D.P.X., a standard microscopical mounting agent, on to a $\frac{1}{4}$ -in. patent plate glass.

The initial diffraction patterns were formed by allowing light to diverge from the star image of the source through the object, and thus on to the plate. The effective positions of source and object were ascertained by using the Aero-Ektar to form images of each, which were located by the eyepiece. The actual positions were then obtained by Newton's formula. At a later stage these observations were correlated with the index positions of the holders, and a "zero error" established for each.

Owing to the readiness with which quarter plates can be obtained, these were taken as a standard plate size. A $2\frac{1}{4}$ -in. square aperture at one end of the plate received the diffraction pattern, through a hole in the mask, and an area $2\frac{1}{4}$ in. \times $\frac{1}{4}$ in. at the other end was subsequently exposed to the same radiation through a Kodak No. 1 step wedge. This consists of a set of densities from 0 to 1.5 in steps of 0.15 approximately. It also carries three coloured patches, red, green and blue, originally intended for the identification of three-colour separation negatives. To modify it to identify the four principal mercury lines, a small patch of Wratten 22 was added. This step wedge enables a measure of control to be exercised over the processing, and also provides an index of the radiation by the use of the patches.

The plate is loaded into a printing frame, carrying the mask, in the dark-room and wrapped in black paper for transfer to the optical bench. It is located in the holder, and exposure is effected by removing the black paper for the required time. The step wedge is exposed immediately afterwards.

Development is carried out in an ordinary M.Q. developer to a medium contrast, depending on the emulsion used. An estimate of the contrast is made by eye, and the printing technique varied accordingly. Low contrast negatives are printed on process plates (by contact) and high contrast negatives on "Ilford Ordinary" which gives a softer result. In accordance with the recommendation of Gabor an over-all contrast of

about two is aimed at, but it is not found necessary to achieve this with any very great precision. A value a little less, if anything, is favoured, as higher values produce a characteristic effect of "burning out" dark lines, to give an artificially light centre.

The positive print is mounted with D.P.X. on a $\frac{1}{4}$ -in. plate, as for the object, and placed in the hologram holder. The object is removed, and the reconstructed image is photographed directly with the plate holder, or indirectly with the Areo-Ektar and lensless camera. The former is generally preferred. Possible arrangements are shown in fig. 9.

§ 7. EXISTENCE OF A FOCUS

As indicated in the theoretical sections, significance is attached to the parameter, f , and experiments were directed to the elucidation and verification of this idea and its consequences.

With the H.P. Lamp, the major part of the work was done with star images. In particular, a hologram negative (H 138) was taken in blue light (4358) of an object which was a photographic replica (by contact printing) of a micrometer eyepiece scale 1 cm. long divided into 100 parts.

A positive print of this (H 140) was mounted and subjected to careful study in Hg green light. In particular the source-hologram distance was varied and the image located. There are two ways of doing this, both using an auxiliary lens. In the first place the auxiliary lens is used to form a real image of both source and subsidiary images. These may be located by an eyepiece. The method of no-parallax is *not* available since all the light passes through a single point region, and no differences of viewpoint are available. Location must thus depend on an impression of sharpness by no means easy to ascertain with a highly coherent illuminating system. It was thus felt that a single eyepiece determination would be of doubtful value, and in practice a series of at least six observations, with the auxiliary lens moved between, were taken as a single "location run" and the image position calculated from each observation. The mean of these observations was taken as the position of the image.

This procedure, though tolerably accurate, was tedious and fatiguing. The following alternative was developed. The auxiliary lens was used to produce a real image of source and reconstructed images, and also of the object, left in the path between the source and hologram. Normally either object or hologram is inverted to avoid overlapping. If, now, the object is moved back and forth, it may be brought into coincidence with one (but only one) of the reconstructed images, as judged by the fact that object and reconstructed image are equally sharp as seen in the

eyepiece. A photograph of this composite field is shown in fig. 10, Pl. II.

In this way it was only necessary to carry out one location run, at the beginning of the series, and get all other positions by observing the

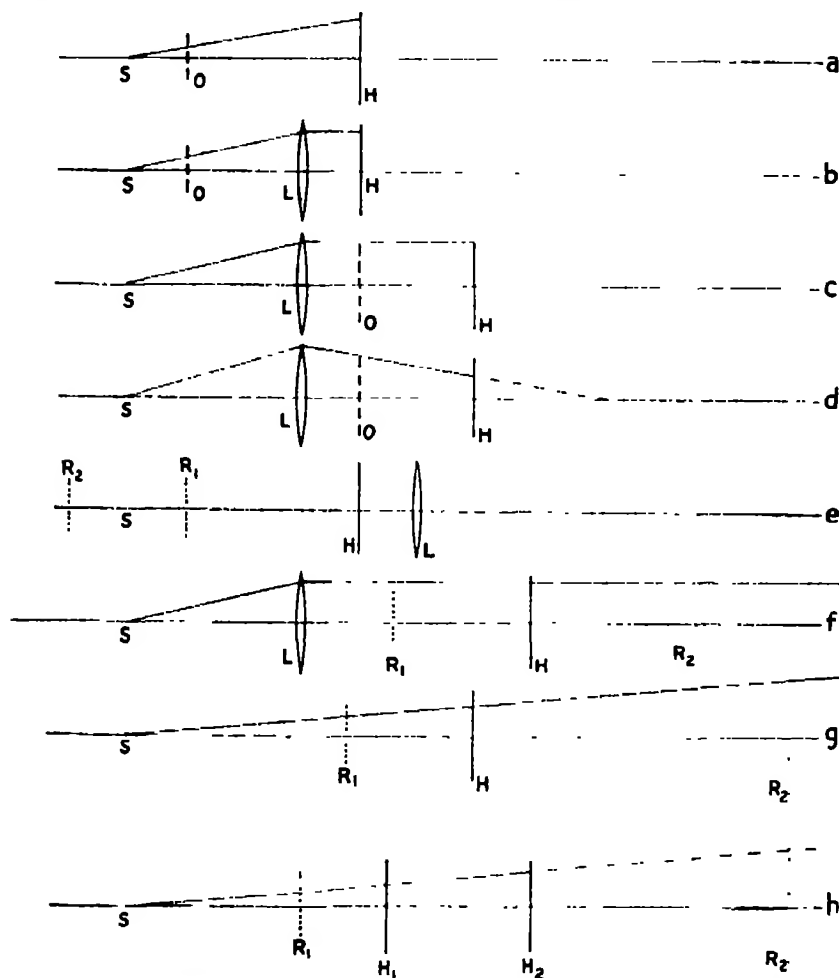


FIG. 9.—Some suggested systems. ——— object, ——— hologram, reconstructed image. (a) Direct divergence. (b) Indirect divergence. (c) Parallel light. (d) Convergent light. (e) Gabor's original reconstructing system. (f) Reconstruction in parallel light. (g) Lensless reconstruction in divergent light. (h) Set-up for hologram of a hologram (H 157/9)

shifts of arbitrary indices fixed to the object and hologram carriages. At the end of the operation, a new location run was performed as a check.

Furthermore, since the object was a scale, and the hologram was

produced from it and gives a scale on reconstruction, the effective over-all magnification of the operation can be determined at each stage by the direct comparison of the two scales, real and reconstructed, a process which gives valuable additional information as to the nature of the operation.

In this operation a slight difficulty arises in locating object and hologram from the fact that they are mounted between $\frac{1}{2}$ -in. glass plates, and hence their optical position differs from their physical position. The method chosen is to locate them using the auxiliary lens, and to calculate their distance from a known focal point of this lens, using Newton's formula, on the assumption that the medium is air. This automatically gives their effective optical position. There is, however, some slight suggestion that the effective optical position varies slightly with the angle of the cone accepted by the lens.

Table I gives the index readings and located positions of source, scale and hologram at the beginning and end of the run.

TABLE I

	Source	Scale		Hologram		Zero Errors	
	Located	Index	Located	Index	Located	Scale	Hologram
Beginning	95.75 cm.	94.0	94.10	87.88	85.34	.10	-2.54
End	95.64 ..	91.0	91.07	79.85	77.08	.07	-2.77

It is not altogether clear whether the change is due to errors in location, or whether they arise from a slight drift due to change in the effective optical position. Calculations have been made on both bases, those on the second assumption being given in Table II.

Positions are estimated to $1/10$ mm., even when location is less accurate, to ensure that errors arising from the use of rounded figures in calculation shall be smaller than those due to the observations. It will be seen that both the focal length (or the power $1/f$) and the expression $Mu/(u-v)$, are sensibly constant. Experience shows that errors of 2 mm. in setting the hologram can easily occur, and hence an accuracy better than ± 5 per cent. cannot be guaranteed.

The constancy of the power was expected on theoretical grounds. The significance of the expression $Mu/(u-v)$ now requires exploration. Gabor has shown that, in the production of the hologram, the coarse structure, which is reproduced directly like a shadow, is magnified by the usual projective law. A similar law is expected to hold for the reconstruction of the image. Since the coarse structure must act as a species of "framework", it is legitimate to expect that the fine structure, when ultimately reconstituted, will also be found to obey the same law. This has been verified by experiment.

The over-all magnification M of the whole process, as observed by direct comparison of the two scales, is recorded in the 11th column. Now the projective magnification of the 2nd or reconstructive stage will be less than unity, and is given by $(u-v)/u$ in our notation. The reconstructed scale is smaller than the coarse framework recorded on the hologram. Since we know the observed over-all magnification, and also the theoretical stage 2 magnification (column 13), it is legitimate to divide the first by the second to get the effective magnification at the first stage. That is, the magnification between the original object and the coarse framework on the hologram. Since this is fixed once and for all by the initial exposure conditions, we anticipate that $Mu/(u-v)$, its measure, will also be constant, and the constancy of this expression constitutes the verification of the theory.

It is also found on reference to the conditions under which H 138 was taken, that the source was at 95.84, the scale at 93.04, and the plate at 80.0. From this we deduce a projective magnification of $\frac{15.84}{2.79} = 5.65$ in excellent agreement with the value of Table II (last column). In fact, allowing for the probable errors of location, a variation of 5 per cent. would not have been excessive.

Now we note that H 138 was taken in blue light, but the reconstructions of Table II were made in green light. We see from this that the coarse structure is magnified in a purely projective way and the resultant magnification can be calculated on this basis from source, object, hologram, and reconstructed image positions without *direct* reference to the wavelengths involved. Of course, the wavelength used has an indirect influence on the position of the reconstructed image, by its influence on the effective hologram focal length. Moreover, we are here concerned with the linear magnification: we are not concerned with the resolving power and some of the magnification may be "empty". The resolving power is very closely related to the wavelength.

The data for the conditions under which H 138 were taken lead to a calculated hologram power of 1.36 dioptries in blue light. This, of course, varies with wavelength, and so in green light we expect a power of $1.36 \times \frac{5461}{4358} = 1.704$ dioptries. The errors in the positions of source and scale as located prior to taking H 138 could produce a variation of ± 5 per cent. in the power, which is sufficient to bridge the gap between this and the value in Table II (Column 10).

§ 8. WAVELENGTH VARIATIONS

The above results, while not inconsistent with the theory that the power varies directly as the wavelength, cannot be regarded as a particu-

larly convincing demonstration of the fact. To investigate this phenomenon, the following experiment was performed.

Two holograms were to be taken with the same source, object and plate positions, but in different wavelengths. Bromide enlargements were to be made from each negative under fixed conditions of magnification. These enlargements were then to be cut along a line, and a "top" of one mounted above a "bottom" of the other. The enlargements were then photographed down to approximately the original size and a positive composite hologram was prepared from them.

Two attempts have been made along these lines. The first was an early attempt with the Osira. Only two wavelengths with this were sufficiently intense for use in this experiment, the blue and the green. They do not have a very large ratio (1.25:1) and hence were not very promising. H 63 was given an hour's exposure to green light (being an orthochromatic plate), while H 65 was the same object, in blue light using a fast ordinary plate for about 10 minutes. The composite hologram produced from this was called H 68, and was studied intensively in green light. The upper half of the hologram was from H 65, and the lower from H 63, the object being a reduced copy of a newspaper cutting with a chess problem.

The second experiment was done with the H.P. arc. The set-up was maintained accurately by clamping down the new object holder and plate holder to the bench. A series of photos was taken in blue and in red light, thereby getting a more favourable wavelength ratio of ~ 1.42 . Two were selected as being of comparable density and contrast, as judged from the hologram and step wedge respectively, so as to get bromide enlargements more closely matched than before. In the end H 186 and H 187 were chosen.

H 186 was 1 hour's exposure in red light to a fast panchromatic plate (P 1500), the object being a microscope eyepiece scale. H 187 was 2 minutes' exposure on a slow ordinary plate (Ilford Ordinary). After enlarging, the holograms were trimmed to show the scales and a small surround, and were mounted with the scales opposite one another, one being inverted for the purpose. An enlargement to the same degree was made of the step wedges, which were carefully matched and included on the hologram to enable the over-all contrast to be checked. On re-photographing the resultant positive was numbered H 193. In this case great care was taken to get the final hologram the same size as the originals, the point being checked on the focusing screen with one of the original negatives. This is to avoid any complications with the scale factor.

A run was not performed with H 193, but instead observations were made in green light rendered parallel by the auxiliary lens. In this case the focal length can be determined directly as the distance the eye-

TABLE II.—Experiments in Diffraction Microscopy
(Mercury green light $\lambda = 5461 \text{ \AA.}$)

Source	Scale		Hologram		(cm.)		(Dioptres)			Mag. overall observed	$u - v$		$\frac{Mu}{u - v}$
	Index	Position	Index	Position	"u"	"v"	$1/u$	$1/v$	$1/f$		cm.	$\frac{u - v}{u}$	
96.75	94.0	94.10	87.88	85.34	10.41	8.76	9.607	11.416	1.809	0.903	1.65	-1596	5.695
95.74	93.7	93.80	86.95	84.39	11.35	9.41	8.811	10.627	1.816	0.970	1.94	-1710	5.670
95.73	93.4	93.50	85.96	83.37	12.36	10.13	8.091	9.872	1.781	1.007	2.23	-1804	5.670
95.72	93.1	93.19	85.05	82.44	13.28	10.75	7.530	9.303	1.773	1.082	2.53	-1905	5.665
95.71	92.8	92.89	84.60	81.97	13.74	10.92	7.278	9.157	1.879	1.150	2.82	-2052	5.610
95.69	92.5	92.59	83.70	81.05	14.64	11.54	6.830	8.665	1.835	1.188	3.10	-2117	5.615
95.68	92.2	92.28	82.30	79.62	16.06	12.66	6.227	7.899	1.872	1.210	3.40	-2119	5.715
95.67	91.9	91.98	82.00	79.30	16.37	12.68	6.109	7.886	1.777	1.250	3.69	-2256	5.640
95.66	91.6	91.68	81.20	78.48	17.18	13.20	5.821	7.576	1.755	1.299	3.98	-2317	5.606
95.65	91.3	91.37	80.30	77.55	18.10	13.82	5.525	7.236	1.711	1.321	4.28	-2368	5.585
95.64	91.0	91.07	79.85	77.08	18.56	13.99	5.388	7.148	1.760	1.356	4.57	-2465	5.505

Av. 5.616

Av. 1.779

TABLE III.—H 68 Hg Green Light. Examination of "Symmetrical" Images

Hologram "u"	$1/u$	U_1 Holog. "v ₁ "	$1/v_1$	L_1 Holog. "v ₁ "	$1/v_2$	$1/f_1$	Average	$1/f_2$	Average
11.19	8.94	14.38	6.96	13.46	7.43	1.98	1.99 ₀	1.51	1.51
12.72	7.87	17.08	5.85	15.71	6.37	2.02		1.50	
13.56	7.38	18.59	5.38	16.99	5.88	2.00		1.49	
14.75	6.78	20.62	4.85	18.92	5.29	1.93		1.47	
15.54	6.43	22.52	4.44	20.18	4.96	1.99	2.05	1.60	
16.64	6.01	25.26	3.96	22.69	4.41	2.05			
				U_2 Holog. "v ₂ "	$1/v_3$	$1/f_1$		$1/f_2$	
11.19	8.94	9.64	10.38	8.93	11.19	1.44	1.39	2.25	2.01
12.72	7.87	10.84	9.22	10.20	9.80	1.35		1.93	
13.56	7.38	11.40	8.77	10.68	9.37	1.39		1.99	
14.75	6.78	12.33	8.10	11.41	8.77	1.32		1.97	
15.54	6.43	12.72	7.86	11.91	8.40	1.43	1.40	1.97	
16.64	6.01	13.59	7.41	12.60	7.93	1.40		1.92	

piece moves between focusing on the hologram and on the reconstructed image. A number of settings were made, which showed variations of ± 3 mm. for the shorter focus, and ± 6 mm. for the longer focus. Both are thus within $\pm 2\frac{1}{2}$ per cent. The focal lengths were 20.31 cm. and 28.91 cm., corresponding to powers of 4.93 and 3.46 dioptries respectively. The ratio of these powers is 1.425, in excellent agreement with the ratio of 1.42 of the wavelengths.

A further check is obtained from the taking set-up. The spacing of source, scale, and plate leads to a power of 3.87 dioptries; this being 3.87 dioptries in blue for H 187 and in red for H 186. When these powers are converted from blue and red into green they come out at 4.84, and 3.41 respectively, which gives good agreement with the observed powers.

It seems fair to accept a direct relation between power and wavelength, as indicated by the theory. This being so, we can go a stage further, and get a more fundamental constant of the hologram. Hitherto we have specified the power of a given hologram in a given wavelength, but since these are proportional to one another, we can now specify the power per unit wavelength. In order to get numerically convenient quantities from ordinary holograms it is suggested that the power per unit wavelength be normally specified in dioptries per micron. We get for H 187, then, a power per unit wavelength of 8.88 dioptries/micron and for H 186 of 6.25 dioptries/micron. It will be observed that this unit has the dimensions L^{-2} and a magnitude of 1/sq. mm. It is called the power-rate.

For comparison it might be pointed out that for an electron system with a source-object distance of 0.1 mm. and a source-plate distance of 1 metre, using electrons of 0.05 Å. wavelength, the power-rate will be about 20 dioptries/micron. Such a hologram will thus be comparable with the optical ones here considered.

§ 9. THE SCALE FACTOR

It was early appreciated that the scale on which the hologram positive actually used is reproduced from the negative as originally taken, would be important and would be worth study as throwing light on the theory of the process. While, therefore, contact printing was normally used for reproduction, an enlarger was also employed in some cases.

In the first place a couple of pairs of holograms were prepared: each containing (a) a contact print, and (b) a *reduction* in the enlarger. The latter was preferred to an enlargement, because it gave a hologram easier to handle. This arises from the fact that such a hologram has an increased power-rate.

In one case the pair was examined in parallel light, and the focal lengths determined as for the composite hologram, H 193. We got, for

the normal size, $f = 21.47 \pm .59$ cm. and for the reduced hologram $f = 11.35 \pm .34$ cm. This leads to a ratio of $1.89 \pm .11$. Careful measurements on these holograms give the linear ratio as $1.351 \pm .01$. This corresponds to a (ratio)² of $1.8252 \pm .03$. It will be seen that the law $f \propto L^2$ holds within the limits of experimental error.

The other pair was measured without an auxiliary lens, by u and v measurements over a considerable length of bench. The powers obtained were: Normal 0.55 dioptre, reduced 1.46 dioptre. Ratio 2.67. Linear reduction 1.573, which squared gives 2.48.

As a further check, one particular hologram was enlarged. This was H 175, taken in parallel light, with the unusually high power-rate of 17.3 dioptries/micron. This is the only hologram which we felt strong enough to stand "dilution". It was subjected to a linear magnification of $2.76 \pm .07$ and focal lengths were obtained as follows (in Hg green). Normal $10.61 \pm .2$ cm. Enlarged 78.92 ± 1 cm. Ratio $7.44 \pm .24$. Square of linear magnification $7.62 \pm .35$. Here again, the agreement is satisfactory, and the fact that the linear ratio here departs substantially from unity is an additional check on the theory.

§ 10. THE TWO IMAGES

It is an essential of the theory that both zone plate and hologram produce two images. So far we have only examined one image at a time. The experiments first described located the image produced by H 140 acting as a divergent lens. Some of the later experiments, determining the focal length directly, use the hologram as a converging lens. It remains to show that the converging and diverging powers are the same.

Reference may here be made to some work with the early composite hologram H 68 in green light. The work was all done with the old Osira and was thus not carried out under the most favourable conditions. The method of locating an image with a moveable object had not been developed (and would not have located the second image). So all data were obtained from "Location runs". This makes the analysis very tedious, and the work has not been repeated.

Measurements were made to locate (i) the upper image sharp, (ii) the lower image sharp, (iii) the source, (iv) the "symmetrical" lower image sharp, and (v) the "symmetrical" upper image sharp. The hologram itself was located separately.

It at once became apparent that the "symmetrical" images were not symmetrical. Gabor's deduction of symmetrical images rests on the assumption that the image source distance is very small compared with the source-hologram distance, *i.e.*, that f is very large. In this case, indeed, the lens formula does give symmetrical images.

The results are summarized in Table III.

An examination of Table III shows at once that the results obtained from the upper part of the hologram (U_1 and U_2) are as satisfactory as can be desired. This part of the hologram contained resolvable writing, and was processed to very nearly the ideal contrast. Hence this image was easy to locate. The lower part, L_1 and L_2 , does not give the same sort of agreement. Here the hologram contains no resolvable writing, and is of lower contrast and hence the locations were more difficult. It is not felt that the divergence is beyond the limits of experimental error.

Working from the estimated power on taking, and assuming the zone-plate law, we get theoretical powers of 1.67 and 2.01 dioptres for the two cases. But owing to the practical difficulties of being sure of the correct positions before the taking device was made, these figures must be taken as indicating an order of magnitude only.

We see, therefore, that the theoretical expressions derived from the simple theory of the zone plate are found experimentally to hold for holograms in general, within the limits of experimental error. We therefore feel justified in concluding that the hologram is a generalized zone plate and in particular we associate it with the following properties:

(1) A given Fresnel diffraction pattern, produced in coherent monochromatic light of wavelength λ proceeding through a point may be associated with a focal length f , being the focal length of that lens which, placed in the plane of the pattern, images the source (or virtual source) in the plane of the object producing the pattern.

(2) If a photographic reproduction of the pattern is now placed in a coherent beam of light of wavelength λ it will form two "images" of the source, resembling the object, as though it were a lens of focus $\pm f$.

(3) If it be placed in a beam of wavelength λ' it will act as if it had a new focal length $\pm f'$ where $f'\lambda' = f\lambda =$ a constant of the pattern. In particular $1/f\lambda$ is called the power-rate.

(4) If the scale of the pattern be altered by a linear factor L , it will be found that the constant factor $f\lambda$ is altered by a factor L^2 .

(5) The effective magnification between the original object and its reconstruction can be calculated by purely projective considerations, from the original source through the object to the hologram, and then from the hologram through the reconstruction to the reproducing source.

The projective law follows from the "coarse structure" relation, and means that the hologram of an object (as against a point or line) must be described by *two* parameters. One is f : this gives the size of the fringe structure round any particular point. The second is the projective magnification, this determines the separation between the *centres* of these fringe systems in terms of the separation of the two original points.

This also explains why "high order" images are not normally observed. If we reduce f to $f/3$ we will normally reduce the fine structure in one ratio and the coarse in another: one varies with f and the other with the projective magnification.

§ 11. MULTIPLE OPERATIONS

When it has been proved that an operation can be performed once, it is frequently instructive to perform it twice. Accordingly H 140 was taken and put about a metre (103 cm.) from the source in green light. Under these conditions it would produce two images, a real image about 123 cm. further away in a direction opposite to the source, and a virtual image 36 cm. on the source side of the hologram. A plate was put up 45 cm. on the more distant side of the hologram, so that the light passed through the hologram and fell on the plate before forming the real image. The plate thus records, in effect, two images, one being 81 cm. one side of it, and the other (unformed) 78 cm. on the other side of it.

The hologram, H 159, formed from this —ve has *two* effective focal lengths, $\pm f_1$ and $\pm f_2$, and produces *four* images from a point source. It is shown in Fig. 15, Pl. II.

Consideration of the taking arrangement suggests that these two foci should be ± 51.2 cm. and ± 180.3 cm., but that neither is likely to be estimated to better than ± 5 per cent., if that. Direct measurement of the foci gives 47.1 and 168.6 cm. respectively, again to ± 5 per cent. or a little better. In view of the fact that none of these images is as easy to see as is the case with a simple hologram, this agreement may be regarded as satisfactory.

Now one method of reconstruction is virtually this: the hologram is put, say, 100 cm. from the source and forms a real image at a greater distance, say 223 cm., as above. If a plate be placed here, it will record the image and this constitutes a useful and legitimate method of reconstruction. But this reconstruction is itself a hologram of a hologram and should have two focal lengths. One will be zero: and as $+0$ is the same as -0 we have here a case of coincident roots. The hologram and image become the same thing. But there will also be a subsidiary pattern from the virtual image between the source and the original hologram. Gabor recognizes this when he says that the unwanted image interferes to a slight extent with the wanted image. But the unwanted image does more than that; it records on the reconstruction, albeit diffusely, *all the information required to reconstruct it*.

We took one of our reconstruction negatives (R7) (fig. 12, Pl. I) and printed it to form a hologram H 162. This hologram was examined in parallel light, and as well as having the trivial focal length ± 0 (giving a

sharp image in its own plane) it was found to have a focal length of 141.0 cm. within the usual limits. The anticipated focal length from the details of R7's taking was 145.5 cm., in satisfactory agreement.

We also made a reconstruction (R 11) (fig. 14, Pl. I) from this hologram to provide visual proof of the existence of the secondary image associated with the original reconstruction R 7.

Of course, this can in theory be done indefinitely. R 11 is a third-order hologram, and in theory might give rise to eight images. In practice owing to the existence of several coincident roots, there will be five, one of which coincides with its own plane. No other third-order holograms have yet been produced, but it might be instructive to try and produce one from H 159 in such a way as to produce coincident roots in a place other than in its plane. This might give an extra clean image, or even give rise to "beating" effects in the plane of the third-order hologram.*

§ 12. MISCELLANEOUS EXPERIMENTS

It is inevitable in an investigation of this kind that a number of side experiments are set up to explore minor ramifications of the process. A number of these are worth putting on record.

(a) An early experiment was made with two objects so sandwiched in glass as to lie in different planes, and thus at different distances from source and hologram. The hologram so produced, reconstructs the two objects in different planes, with such overlapping of the out-of-focus outlines as would occur in direct viewing through an equivalent system. The phenomenon can be explained by noting that each contributes to the net effect a diffraction pattern with a characteristic (but different) focal length, and the positions of the reconstructed images can be obtained from this.

(b) A similar effect can be obtained by taking two holograms of different objects with different power-rates. These are then double-printed on to a single plate, i.e., a given plate is exposed an appropriate time behind each negative in turn and then developed. The diffraction pattern which results differs in detail from that of case (a) but the visual effects which result are very similar. Once again the two images can be partially separated by their different focal lengths.

(c) In one case a copy negative was made from a positive (with an over-all γ from the original pattern of ~ 2) and used as a hologram. This has the same focal length as the positive, and produces a reconstructed image which is a negative. R 7 and R 8 (figs. 12 and 13, Pl. I) are reconstructions from a positive and a negative hologram (positive

* Note added 5.9.51. This has since been identified as a reconstruction of the first order hologram.

hologram shown as fig. 11, Pl. I) under identical conditions, and they are mutually super-posable to give a sensibly uniform density. The exact theory behind this is not known, as it was not expected from Gabor's analysis, but the exactitude with which it works is impressive. If the initial negative can be developed to a γ of 2 (and this will only be possible with certain emulsions) there is a possibility here of economy in running costs. For a plate or paper set up in the reconstruction plane now gives a direct + ve. As the plates used here are not all suitable for high γ working (and this has other disadvantages) the method has not been much used, but some of the electron-sensitive emulsions will be more amenable to this type of treatment.

(d) It follows from (b) and (c) that holograms can be subtracted. In particular, it should be possible to subtract a hologram of the background from that of background + wanted image, to obtain the latter alone. The successful accomplishment of this would require a carefully adjusted "dummy" object (plain glass of equivalent thickness) and hence it has not been attempted.

(e) A relief image in plain (transparent) gelatine has been prepared by the Carbro process from a hologram (H 135) of the scale. By itself, this relief image introduces excessive phase distortions into a coherent beam. But if a covering plate is mounted over it using D.P.X. cement, there remains just sufficient difference of refractive index between the gelatine and the D.P.X. to produce a delicately graded phase-contrast object. This phase contrast hologram reconstructs as before. The analogy is with a zone-plate in bichromated gelatine with a $\frac{1}{2}\lambda$ retardation in the gelatine areas. As is well-known, this zone-plate acts like a normal one, but with greater light-gathering power.

(f) A bromide enlargement was made from a hologram negative, and the dark fringes were inked over with Indian Ink. The bromide image was then removed with Farmer's reducer to leave the ink drawing on an otherwise white ground. This was photographed down to form a hologram (called the black-and-white hologram). The subject was the microscope eyepiece scale, and the black and white hologram, though only containing three or four fringes, gave a crude but recognizable reconstruction. The method is chiefly advantageous in allowing background dirt to be ignored. A fuller description, with photographs, is given in Rogers (1950).

(g) An attempt was made to produce a composite hologram containing two side-by-side images in dye, produced by dye-coupling developers. The idea was to use the whole radiation to illuminate the system and get one side of the hologram absorbing only one-wavelength, the other only another, and thus effecting reconstructions in two different wavelengths in adjacent positions. But the dyes used did not interact in the way hoped, and tended to absorb too wide a wavelength band.

(h) An attempt was made to produce a hologram in a beam of parallel light and this was successful. The only conditions are that the object shall be small compared with the diameter of the beam of parallel light, and that the hologram shall be recorded reasonably near to the object, so that the diffraction pattern of the object still falls wholly within the area of plate illuminated by the parallel beam. The object-plate distance is automatically the focal length of the hologram, and this is how the short-focus (high power) hologram mentioned in § 9, was obtained.

(i) It is worth placing on record the fact that we have, in effect, produced a hologram in convergent light. About four years ago, while doing some experiments with the honours class a number of Fresnel diffraction patterns were produced in convergent light. The light from a small source (Hg Arc with pin-hole) was made slightly convergent with a telescope objective. It was found in this way that (i) a slightly larger pattern could be condensed on to a slightly smaller photographic plate, and that (ii) a concentration of light was produced which gave a welcome reduction of exposure.*

§ 13. METHODS OF RECONSTRUCTION

It is convenient here to summarize the methods of producing a reconstruction from a hologram. (i) There is the original Gabor method of projecting the small reconstructed image near the original object position, on to a plate by the use of an auxiliary lens. (ii) The light from the point source may be rendered parallel by an auxiliary lens, and the image observed or photographed in the focal plane of the hologram. (iii) The auxiliary lens may be dispensed with, and the divergent beam passed through the hologram, which must be further from the source than its own focal length, and the resultant real image observed or photographed.

Methods (ii) and (iii) are recommended, as avoiding very fine-grain techniques in the photography. Method (ii) conserves the light well over long distances, and hence exposures are kept low. On the other hand the last method allows further projective magnification to be obtained, at some cost in exposure time, if desired.

We understand that Gabor (Gabor, 1950) has discovered method (ii) independently.

§ 14. ARTIFICIAL HOLOGRAMS

Some work has been done on artificial (calculated) holograms, including linear zone-plates, but a discussion of these experiments is reserved for a further paper.

* Note added 5.9.51. This has since been confirmed by reconstruction.



FIG. 11.—H 146, from a microscope eyepiece graticule



FIG. 12.—R 7, a reconstruction from a positive of H 146

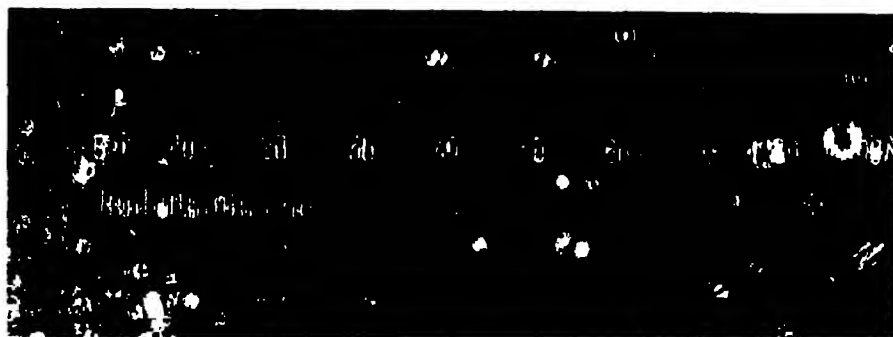


FIG. 13.—R 8, a reconstruction from a negative of H 146

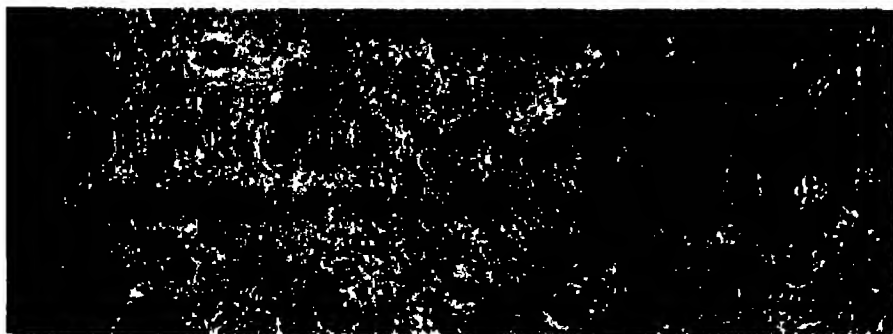


FIG. 14. R 11, a reconstruction from a positive of R 7
(a third order hologram)

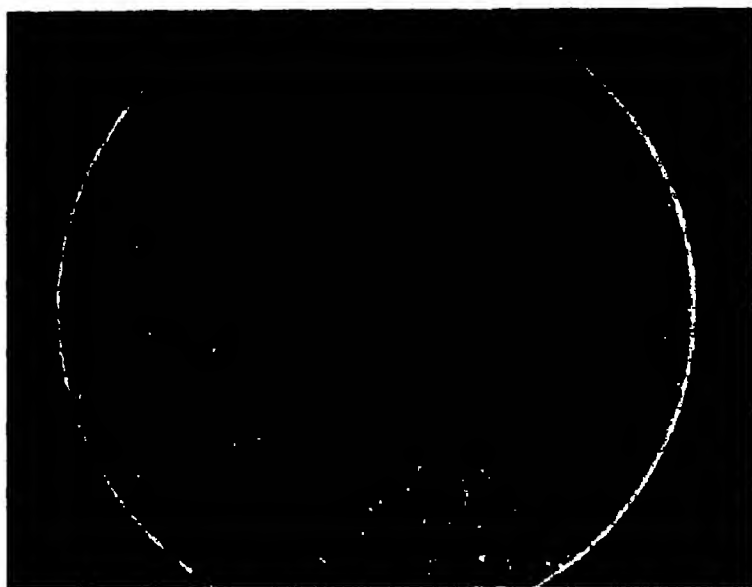


FIG. 10.—Composite image consisting of a reconstruction (above) alongside an image of the object (below), located in the plane of reconstruction, as used for the work of para. 7, using a Gabor system (fig. 9e)



FIG. 15.—H 159, a hologram of a hologram

§ 15. ACKNOWLEDGMENTS

My thanks are due to Professor G. D. Preston for his current help and encouragement; To Professor Gabor for coming to Dundee and giving a lecture which sowed the seeds of many of these ideas, including the zone-plate analogy in its simplest form, and for periodic encouragement since; to Messrs. Pilkington Bros. for a supply of their best $\frac{1}{4}$ -in. plate glass, without which the work would have been prohibitively expensive; to Professor A. C. Lendrum and the D.R.I. for supplies of D.P.X.; to Mr. A. Gunn for advice on the Carbro process; to Dr. C. Waller, of Ilford Ltd., for preparing me a batch of specially sensitized Ilford H.R. plates; to the workshop staff and especially J. and W. Stark for their patience with my mechanically exacting demands.

REFERENCES TO LITERATURE

- AIRY, 1877. *Undulatory Theory of Optics*, Macmillan, 1877, pp. 60-68.
BRAGG, 1950. *Nature*, CLXVI, 399.
GABOR, 1948. *Ibid.*, CLXI, 777.
—, 1949. *Proc. Roy. Soc. (A)*, CXCIV, 454.
—, 1950. *Private Communication*.
JENKINS and WHITE, 1937. *Fundamentals of Physical Optics*, McGraw Hill, pp. 184-201.
ROGERS, 1948. *Journ. Sci. Instr.*, XXV, 284.
—, 1950. *Nature*, CLXVI, 1,027.
STRONG, 1938. *Modern Physical Laboratory Practice*, Blackie, p. 69, footnote.

(Issued separately February 9, 1952)

XV.—Theorems on the Convergence and Asymptotic Validity of Abel's Series.* By A. J. Macintyre and Sheila Scott Macintyre, University of Aberdeen

(MS. received March 3, 1951. Read November 5, 1951)

SYNOPSIS

In this paper we discuss the Abel series for a function $F(z)$ which is regular in an angle $|\arg z| < \alpha$ and at the origin. We investigate conditions under which the series converges and conditions under which its sum is asymptotically equivalent to the function $F(z)$ in the half-plane $R(z) > 0$.

1. INTRODUCTION

Associated with a function $F(x)$ defined and infinitely differentiable for $x \geq 0$ is its Abel series (Abel, 1839)

$$\sum_{n=0}^{\infty} z(z-n)^{n-1} F^{(n)}(n)/n! \quad . \quad . \quad . \quad (1)$$

If this series converges for a single value of $z \neq 0$, it converges uniformly for each bounded region of the z -plane and its sum is an integral function of exponential type (Halphen, 1881). Abel was possibly not aware of this property as he applied the series without comment in a note which was published posthumously (Abel, 1839, page 82) to the function $\log(1+z)$. Halphen investigated the series for $\log(1+z)$ in considerable detail (Halphen, 1881) with the aim of showing that it could not represent $\log(1+z)$. Much subsequent work has been devoted to the convergence and summability of the Abel series to the function $F(z)$ when $F(z)$ is an integral function satisfying certain conditions (Gontcharoff, 1935; Gelfond, 1938; Buck, 1948).

The present paper supposes $F(z)$ to be regular in an angle $|\arg z| \leq \alpha$ and investigates conditions under which its Abel series converges. The sum of the series is not generally equal to $F(z)$, but we show that under slightly more stringent conditions it has an asymptotic validity in the sense that the difference between $F(z)$ and the series tends to zero as $R(z)$ tends to infinity by positive values. Our principal results are

Theorem 1: *If for $|\arg z| \leq \frac{1}{2}\pi$, $F(z)$ is regular and satisfies*

$$|F(re^{i\theta})| \leq Kr^{-\epsilon} e^{r^{\delta}(\theta)} \quad . \quad . \quad . \quad (2)$$

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

for some positive K and ϵ then the Abel series for $F(z)$ converges uniformly for z in any bounded domain.

and

Theorem 3: If for $|\arg z| \leq \frac{1}{2}\pi$ and $R(z) \geq -h$ ($h > 0$), $F(z)$ is regular and satisfies

$$|F(re^{i\theta})| \leq Kr - \gamma e^{b(\theta)} \quad (3)$$

for some $K > 0$ and $\gamma > 1$ then representing the sum of its Abel series by $A(z)$

$$|F(z) - A(z)| = O\left(\frac{1}{(x \log x)^\lambda}\right) \quad (4)$$

as $x = R(z)$ tends to infinity by positive values.

In the statement of these theorems, $b(\theta)$ is a certain continuous positive function of θ with period 2π . If $F(z)$ is supposed to be an integral function and to satisfy (3) for all θ then it is a known theorem (Schmidli, 1942) that the series converges to $F(z)$. The proof of Theorem 1 is fairly direct, but before proceeding to the proof of Theorem 3 we have to develop some aspects of the theory of the Laplace transform of functions regular in an angle.

Somewhat more general theorems (Theorems 2 and 4) are also proved. The conclusions of Theorems 1 and 3 continue to hold if certain more drastic inequalities are assumed in a smaller angle.

2. CONVERGENCE CONDITIONS

It is known (Polya and Szego, 1945, III, §§ 116, 265) that

$$1 - \frac{z^n}{n} \leq e^{rb(\theta)} \quad (5)$$

for all $z = re^{i\theta}$ where $b(\theta)$ is the supporting function of the bounded domain B which is represented by

$$Z = we^{1+w} \quad (6)$$

on $|Z| \leq 1$. B is a convex pear-shaped domain containing the origin whose boundary C_w has a vertex at $w = -1$ where the two tangents make angles $\pm \frac{1}{2}\pi$ with the positive direction of the real axis. If we put $z = nJ$ and $w = (\bar{J} - 1)^{-1}$ then C_w is represented in the J -plane by a curve C_J and in the z -plane by a curve C_z whose shape is independent of n . Equality is attained in (5) when z lies on C_z . C_z has a cusp at the origin and its tangents there are the radii $\arg z = \pm \frac{1}{2}\pi$. C_z lies entirely within the sector $|\arg z| < \frac{3\pi}{4}$ except for $z = 0$ and is continuous and rectifiable.

If we now use Cauchy's formula for $F^{(n)}(n)$ taking the contour as C_s , the proof of theorem 1 is immediate. For

$$F^{(n)}(n) = \frac{n!}{2\pi i} \int_{C_s} \frac{F(z) dz}{(z-n)^{n+1}} \quad (7)$$

and from (2) it follows that (with $r = |z|$)

$$\begin{aligned} |F^{(n)}(n)| &\leq \frac{n! K}{2\pi} \int_{C_s} \frac{r^{-\epsilon} e^{r\delta(\theta)} |dz|}{n^{n+1} e^{r\delta(\theta)(1+1/n)}} \\ &\leq K' n! n^{-n-\epsilon} \end{aligned} \quad (8)$$

if we take (as we evidently may, without loss of generality) $\epsilon < 1$, to ensure convergence near $z=0$. From this, the modulus of the n^{th} term of Abel's series satisfies

$$\begin{aligned} \left| \frac{z(z-n)^{n-1}}{n!} F^{(n)}(n) \right| &\leq K' |z| e^{(n-1)|z| \delta(\theta)/n} n^{-1-\epsilon} \\ &\leq K(|z|) n^{-1-\epsilon} \end{aligned} \quad (9)$$

This proves theorem 1.

We obtain a generalization of theorem 1 by taking the contour C'_s consisting of the radii $\arg z = \pm \alpha$ ($0 < \alpha < \frac{1}{2}\pi$) from the origin to the points where they meet C_s and the arc of C_s for which $|\arg z| \leq \alpha$, in place of C_s in (7). Let $\frac{1}{2}\pi \leq \alpha < \frac{3}{2}\pi$, then for $\arg z = \alpha$

$$\left| \frac{z}{n} \right| \log |1 - z/n|$$

increases steadily with $|z|$ from its value $-\cos \alpha$ for $z=0$ to its maximum on C_s (Polya and Szego, 1945, III, § 285). Consequently for z on the radial sections of C'_s we have

$$\left| 1 - \frac{z}{n} \right|^n \geq e^{-|z| \cos \alpha} \quad (10)$$

In order to obtain the inequality (8) from (7) after the contour C_s is replaced by C'_s we must now suppose that $F(z)$ satisfies the inequalities

$$\begin{aligned} |F(re^{i\theta})| &< Kr^{-\epsilon} e^{r\delta(\theta)}, \quad |\theta| < \alpha, \\ |F(re^{\pm i\alpha})| &< Kr^{-\epsilon} e^{-r \cos \alpha} \end{aligned} \quad (11)$$

Thus we have

Theorem 2: *If $F(z)$ is regular for $|\arg z| \leq \alpha$ ($\frac{1}{2}\pi \leq \alpha < \frac{3}{2}\pi$) and satisfies conditions (11) then its Abel series converges uniformly in any bounded domain.*

We may note that on account of the Phragmen-Lindelöf (1908)

inequalities, Theorem 2 implies a more stringent inequality than the first of (11) for some range of θ near α . In fact, we could say that a (truncated) indicator diagram for $F(z)$ is at most B for Theorem 1 and the part of B within $|\arg(z+1)| \leq \alpha - \frac{1}{2}\pi$ for Theorem 2. The argument for Theorem 2 continues to apply when $\frac{1}{2}\pi \leq \alpha \leq \frac{3}{2}\pi$, but the hypothesis then degenerates to $|F(re^{i\theta})| < Kr^{-\epsilon} e^{-r \cos \theta}$ for $|\theta| \leq \alpha$ because of the Phragmen-Lindelöf inequalities. When $\alpha < \frac{1}{2}\pi$, as $|1 - z/n|^{n/|\alpha|}$ has an exponentially small minimum along $\arg z = \pm \alpha$ (on a curve which bears the same relationship to the continuation of C_w as C_s does to C_w), we would have to impose a more stringent inequality still.

3. SOME REMARKS ON THE LAPLACE TRANSFORM

In this section we follow the notation of Polya (1929), and make use of extensions of some of his results to functions which are regular and of exponential type in an angle (Pflüger, 1935-6; Macintyre, 1939; Rabinovic, 1948).

Let K be a closed convex region containing the origin and $h(\phi) = \max_{z \in K} \operatorname{Re}(ze^{-i\phi})$. $h(\phi)$ is the supporting function of K and $x \cos \phi + y \sin \phi = h(\phi)$ is the supporting line T_ϕ of normal direction ϕ of K . K is completely determined by $h(\phi)$. Each boundary point of K has at least one supporting line through it and each supporting line has at least one extreme point in common with K , that is a boundary point which is not an interior point of a segment of a straight line forming part of the boundary of K . Let us denote the set of boundary points of K by C . By $C(\alpha)$ we denote the subset of boundary points through which pass supporting lines T_ϕ of normal direction ϕ with $-\alpha \leq \phi \leq \alpha$. If $\frac{1}{2}\pi < \alpha < \pi$, T_α and $T_{-\alpha}$ will intersect the negative real axis. We denote the points of intersection by D_+ and D_- respectively. They will intersect each other in a point D , say, corresponding to the complex number $z_0 = x_0 + iy_0$. Let the boundary points of K which lie on T_+ and T_- nearest to D be denoted E_+ and E_- . Then the join of K and the triangle DE_+E_- is also a convex set. When $\frac{\pi}{2} < \alpha < \pi$, its supporting function is equal to $h(\phi)$ when $-\alpha \leq \phi \leq \alpha$ and equal to $x_0 \cos \phi + y_0 \sin \phi$ when $\alpha < \phi < 2\pi - \alpha$. We denote this convex set by $K(\alpha)$ and call D the vertex of $K(\alpha)$.

Now suppose that $F(z)$ is regular for $|\arg z| \leq \alpha$ (continuous on the boundary) and satisfies

$$|F(z)| < Ke^{A|z|} \quad . \quad . \quad . \quad (13)$$

The Laplace transform of $F(z)$ defined by

$$f(z) = \int_0^\infty e^{-zt} F(t) dt \quad . \quad . \quad . \quad (14)$$

is regular for $R(z) > A$, and by rotating the contour in (14) from the positive real axis with the radius $\arg t = \theta$ ($-\alpha \leq \theta \leq \alpha$), $f(z)$ is analytically continued into the half plane $R(ze^{i\theta}) > A$ and so into a Riemann surface whose boundary consists of the circular arc $|z| = A$ ($-\alpha \leq \arg z \leq \alpha$), and the two half tangents at the extremities of this arc. When $\frac{1}{2}\pi < \alpha < \pi$ the sector $|\arg(A \sec \alpha - z)| < \alpha - \frac{1}{2}\pi$ appears twice in this surface. It is clear that as $|z| \rightarrow \infty$ in the region of regularity of $f(z)$, we have

$$|f(z)| = O\left(\frac{1}{|z|}\right) \quad . \quad . \quad . \quad (15)$$

uniformly for $|\arg z| \leq \alpha + \frac{1}{2}\pi - \delta$. The inverse transform to (14) is

$$F(z) = \frac{1}{2\pi i} \int_{\Gamma} e^{zJ} f(J) dJ \quad . \quad . \quad . \quad (16)$$

where Γ is a contour consisting, to take a particular form, of a circular arc $|J| = A + \delta$, $|\arg J| \leq \frac{1}{2}\pi + \delta$, together with the half tangents at the extremities extending to infinity so that the real part of J tends to minus infinity (Macintyre, 1939).

With our hypothesis that $F(z)$ is regular for $|\arg z| \leq \alpha$ where $\frac{1}{2}\pi < \alpha < \frac{3}{2}\pi$, (15) is satisfied and the contour can be replaced by that consisting of the negative real axis taken twice in opposite directions and the complete circumference of the circle $|J| = A + \delta$ from $-(A + \delta)$ back to itself. The representation (16) will then be valid for $R(z) > 0$. If now $h(\theta)$ is the supporting function of the convex region K containing the origin and $F(z)$ satisfies the inequality

$$|F(re^{i\theta})| \leq Kr^{-\gamma} \exp\{r h(\theta)\}, \quad -\alpha \leq \theta \leq \alpha, \quad \frac{1}{2}\pi < \alpha < \frac{3}{2}\pi, \quad \gamma > 1 \quad . \quad (17)$$

then the contour Γ can be distorted into the negative real axis from $-\infty$ to $D_{-\infty}$, the supporting line $T_{-\infty}$, the boundary $C(\alpha)$ the supporting line T_{α} and the negative real axis from D_{α} to $-\infty$. The special case in which $F(z)$ is an integral function was considered by S. Schmidli (1942), but his proof applies equally well to functions regular in an angle.

If we also suppose K to be symmetrical with respect to the real axis in order to obtain a simpler enunciation we have

Lemma 1: *If $F(z)$ is regular for $|\arg z| \leq \alpha$, where $\frac{1}{2}\pi < \alpha < \frac{3}{2}\pi$ and satisfies (17) where $h(\theta)$ is the supporting function of a convex region K symmetrical with respect to the real axis then for $R(z) > 0$, $F(z)$ has the representation*

$$F(z) = \frac{1}{2\pi i} \int_{\Gamma} e^{zJ} f(J) dJ \quad . \quad . \quad . \quad (18)$$

where the contour Γ consists of the negative real axis taken twice between

$-\infty$ and the vertex of $K(\alpha)$ and the boundary $\Gamma(\alpha)$ of $K(\alpha)$. $f(z)$ is defined by (14).

The inequality (15) is insufficient for our applications and we have to supplement it. The integral (18) can be replaced by the following :

$$F(z) = \frac{1}{2\pi i} \int_{-\infty}^{-h} e^{zJ} \phi(J) dJ + \frac{1}{2\pi i} \int_{\Gamma(\alpha)} e^{zJ} f(J) dJ . \quad (19)$$

where $\phi(J)$ is the difference between the value of $f(J)$ on the negative axis obtained by analytic continuation from the lower half plane, say $f_-(J)$, and the value obtained by analytic continuation from the upper half plane, say $f_+(J)$. With the notation

$$f(z) = \int_0^{\infty e^{i\phi}} F(t) e^{-zt} dt . \quad (20)$$

to mean that the radius of integration is $\arg t = \phi$, we can evidently use for $\phi(z)$ (z real and < -1), the representation

$$\phi(z) = f_-(z) - f_+(z) = \left\{ \int_{\infty e^{-i\alpha}}^0 F(t) e^{-zt} dt + \int_0^{\infty e^{i\alpha}} F(t) e^{-zt} dt \right\} . \quad (21)$$

Now if $F(z)$ is regular for $R(z) \geq -h$ ($h > 0$), the combined contour in (21) can be distorted near the origin into a segment of the line $R(t) = -h$. From this new representation we see that as z tends to infinity by negative values, the contribution of the integrals along the infinite half-radii is $O\left(\frac{e^{-h\alpha \cos \alpha}}{z}\right)$ while the contribution along the segment of $R(t) = -h$ is $O(e^{h\alpha})$ and hence

$$|\phi(z)| = O(e^{h\alpha}) . \quad (22)$$

Accordingly, we have

Lemma 2: *If $F(z)$ satisfies the hypotheses of Lemma 1 and is regular for $R(z) \geq -h$ for some $h > 0$, $F(z)$ has the representation (19) valid for $R(z) > 0$ and $\phi(z)$ satisfies (22), for z real and negative.*

4. THE ASYMPTOTIC VALIDITY OF ABEL'S SERIES

If we take K to be identical with the convex region B described in 2, so that $h(\theta) = b(\theta)$, and take $\alpha = \frac{1}{2}\pi$, then $K(\alpha)$ is also identical with B . The hypotheses of Lemmas 1 and 2 now coincide with the hypotheses of Theorem 3. At this stage a comparison with Schmidli's analysis seems appropriate. Schmidli (1942) assumes that $F(z)$ is an integral function and satisfies (3) for all z . His conclusion depends on the representation (19) with the integral along the negative real axis missing. In his case the expansion of e^{zJ} in powers of $J e^J$ and integration term by

term leads immediately to his conclusion. We perform the same operation. e^{zJ} is expanded in powers of Je^J as the series

$$\sum_{n=0}^{\infty} \frac{z(z-n)^{n-1} J^n e^{nJ}}{n!} \quad . \quad . \quad . \quad (23)$$

This series is at the same time the Abel series of e^{zJ} regarded as a function of z and a Lagrange expansion of e^{zJ} regarded as a function of J . The resulting series

$$\sum_{n=0}^{\infty} \frac{z(z-n)^{n-1}}{n!} \cdot \frac{1}{2\pi i} \left\{ \int_{-\infty}^{-1} J^n e^{nJ} \phi(J) dJ + \int_{0\infty} J^n e^{nJ} f(J) dJ \right\} \quad . \quad (24)$$

is by Lemma 2 the Abel series of $F(z)$ and will be denoted S .

Now the series (23) converges uniformly to e^{zJ} for J on the boundary of B , but when J lies on the segment $J < -1$ of the real axis the series converges uniformly but not to e^{zJ} . It now converges to $e^{\xi J}$ where $\xi e^{\xi} = Je^J$ and $-1 < \xi < 0$ (Halphen, 1881).

We thus have, since $\phi(J)$ satisfies (22),

$$S = \frac{1}{2\pi i} \int_{-\infty}^{-1} e^{\xi J} \phi(J) dJ + \frac{1}{2\pi i} \int_{0\infty} e^{zJ} f(J) dJ \quad . \quad . \quad (25)$$

and hence, by Lemma 2,

$$S = F(z) - \frac{1}{2\pi i} \int_{-\infty}^{-1} e^{\xi J} \phi(J) dJ - \frac{1}{2\pi i} \int_{-\infty}^{-1} e^{zJ} \phi(J) dJ \quad . \quad (26)$$

A weak form of Theorem 3 is now evident, for if $R(z) = x > 0$ the first integral in (26) is bounded and the second is $O(e^{-x})$. We obtain a better approximation to the first integral as follows. From Lemma 2, the integral in modulus is less than a constant multiple of

$$\int_{-\infty}^{-1} e^{\xi J} e^{zJ} dJ \quad . \quad . \quad . \quad (27)$$

and Theorem 3 will follow from

Lemma 3: *If h is real and positive, then as x tends to infinity by positive values*

$$\int_1^{\infty} e^{-\mu x - \lambda x} d\lambda = O \left\{ \frac{1}{(x \log x)^h} \right\} \quad . \quad . \quad . \quad (28)$$

where

$$\mu e^{-\mu} = \lambda e^{-\lambda} \quad (0 \leq \mu \leq 1 \leq \lambda) \quad . \quad . \quad . \quad (29)$$

To prove the lemma, first observe that since

$$\frac{d\mu}{d\lambda} = - \frac{\mu(\lambda-1)}{\lambda(1-\mu)} \quad . \quad . \quad . \quad (30)$$

it follows that

$$\frac{d\mu}{d\lambda} \rightarrow -1 \text{ as } \lambda \rightarrow 1 \quad . \quad . \quad . \quad (31)$$

It follows from (29) that

$$\mu > \lambda e^{-\lambda} \quad . \quad . \quad . \quad (32)$$

Writing $u = \lambda + \mu$, $v = \lambda - \mu$ (29) becomes

$$u = \frac{v(e^v + 1)}{e^v - 1}$$

Two differentiations of $v(e^v + 1) - 2(e^v - 1)$ suffice to prove that

$$v(e^v + 1) \geq 2(e^v - 1) \quad (v \geq 0)$$

Hence

$$\lambda + \mu \geq 2 \quad . \quad . \quad . \quad (33)$$

Using (33), it is an elementary exercise to show that $\frac{d\mu}{d\lambda}$ is an increasing function of λ . (One differentiation suffices.) Now, using this fact and (32) we see that

$$\frac{d}{d\lambda} (-\mu x - \lambda h) = -x \frac{d\mu}{d\lambda} - h$$

decreases and therefore

$$\frac{d}{d\lambda} (-\mu x - \lambda h) > \log x - h \quad (1 \leq \lambda \leq \log x)$$

and is therefore positive in this range, for large positive x .

We now split up the range of integration and consider first

$$I_1 = \int_1^{\log x} e^{-\mu x - \lambda h} d\lambda.$$

The above analysis shows that the integrand in $I_1 \leq 1/x^{h+1}$. Hence

$$I_1 \leq \frac{\log x}{x^{h+1}} = O \left\{ \frac{1}{(x \log x)^h} \right\} \quad . \quad . \quad . \quad (34)$$

Using (32) again, and making the change of variable

$$\lambda = \log x + \log \log x - v$$

$$\begin{aligned} I_2 &= \int_{\log x}^{\log x + \log \log x} e^{-\mu x - \lambda h} d\lambda \\ &\leq \int_{\log x}^{\log x + \log \log x} e^{-x \lambda e^{-\lambda} - \lambda h} d\lambda \\ &= (x \log x)^{-h} \int_0^{\log \log x} \exp \left\{ - \left(\frac{\log x + \log \log x - v}{\log x} \right) e^v + hv \right\} dv \end{aligned}$$

where μ is defined by (29). (39) is a special case of (26), with appropriate changes in sign, for, writing

$$F(z) = \frac{1}{k-z} \quad (k \text{ real and negative})$$

$$\phi(z) = \int_{\infty e^{-i\alpha}}^0 \frac{e^{-tz}}{k-t} dt + \int_0^{\infty e^{i\alpha}} \frac{e^{-tz}}{k-t} dt$$

$$= -2\pi i e^{-ks}$$

Hence (26) becomes

$$S = \frac{1}{k-z} + \int_{-1}^{-\infty} e^{xt-kJ} dJ - \frac{e^{k-s}}{k-z}.$$

REFERENCES TO LITERATURE

- ABEL, N. H., 1839. *Oeuvres Complètes*, Christiania.
- BUCK, R. C., 1948. "Interpolation Series", *Trans. Amer. Math.*, LXIV (2), 283-298.
- GELFOND, A., 1938. "Interpolation et unicité des fonctions entières", *Rec. Math. (Mat. Sbornik)*, N.S., XLVI, 115-147.
- GONTCHAROFF, W., 1935. "Sur la convergence de la série d'Abel", *Rec. Math. (Mat. Sbornik)*, N.S., XLII, 473-483.
- HALPHEN, G. H., 1881. "Sur une série d'Abel", *Bull. de la Soc. Math.*, x, 1881-1882, 67.
- MACINTYRE, A. J., 1939. "Laplace's Transformation and Integral Functions", *Proc. London Math. Soc.*, (2), XLV, 1-20.
- PFLÜGER, A., 1935-36. "Über eine Interpretation gewisser Konvergenz- und Fortsetzungseigenschaften Dirichletscher Reihen", *Comm. Math. Helv.*, VIII (2), 89-129.
- PHRAGMEN, E., and LINDELÖF, E., 1908. "Sur une extension d'un principe classique de l'analyse", *Acta Math.*, XXXI, 381-406.
- POLYA, G., 1929. "Untersuchungen über Lücken und Singularitäten von Potenzreihen", *Math. Zeits.*, XXIX, 549-640.
- , and SZEGÖ, G., 1945 edition, *Aufgaben und Lehrsätze*, Vol. I, New York.
- RABINOVIC, YU. L., 1948. "Inversion Formulae for two kinds of Laplace Transform", *Doklady Akad. Nauk. S.S.S.R.* (N.S.), LX, 969-972. (Russian.)
- SCHMIDT, S., 1942. "Über gewisse Interpolationsreihen", *Thesis, Zurich*.

XVI.—Some Continuant Determinants arising in Physics and Chemistry—II.* By D. E. Rutherford, D.Sc., Dr.Math.,
United College, University of St. Andrews

(MS. received May 4, 1951. Read June 2, 1951)

SYNOPSIS

Some of the formulae obtained in this paper are likely to find application in problems concerning a rectangular lattice of "atoms", each of which is under the influence of its near neighbours. Some of the determinants considered apply to cases in which both the nearest and the next nearest neighbours are operative. The inverses of certain types of matrices are found, and these may prove to be of value either in solving systems of linear equations such as arise in relaxation problems, or in determining the latent roots of matrices which may occur in problems in applied mathematics.

§ 1. Let $P_n(x)$ denote the square matrix

$$\begin{bmatrix} x & 1 & & & & & \\ 1 & x & 1 & & & & \\ & 1 & x & 1 & & & \\ . & . & . & . & . & . & . \\ & & & & & 1 & x \end{bmatrix}$$

of order n and let $\phi_n(x)$ denote its determinant. This determinant, sometimes called Wolstenholme's determinant, and modifications of it frequently make their appearance in problems in mathematics, physics and chemistry.

In a previous paper the author (1947) evaluated and factorized many determinants allied to $\phi_n(x)$, and certain of these have already proved useful to workers in different branches of science. The present note contains some further results concerning such determinants which, it is hoped, will also prove to be of practical application.

It has been decided to omit detailed proofs of many of the formulae obtained when these depend merely upon trigonometrical and algebraic manipulation; for these formulae, while of interest and use to the physicist and applied mathematician, have much less significance in pure mathematics.

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

We begin by gathering together the more important formulae for $\phi_n(x)$.

$$\begin{aligned}\phi_n(x) &= \prod_{k=1}^n \left(x - 2 \cos \frac{k\pi}{n+1} \right). \quad (1.1) \\ &= \binom{n}{n} x^n - \binom{n-1}{1} x^{n-2} + \binom{n-2}{2} x^{n-4} - \dots\end{aligned}$$

the series terminating either with a constant term or a term in x . Writing

$$x = 2 \cos \theta = \frac{u+v}{\sqrt{uv}}, \quad (1.2)$$

we find that if $|x| \neq 2$, then

$$\phi_n(x) = \frac{\sin(n+1)\theta}{\sin \theta} = \frac{u^{n+1} - v^{n+1}}{u - v} \frac{1}{(uv)^{n/2}}. \quad (1.3)$$

and

$$2 \cos n\theta = \frac{u^n + v^n}{(uv)^{n/2}}.$$

On the other hand,

$$\phi_n(2) = n+1, \quad \phi_n(-2) = (-1)^n(n+1). \quad (1.4)$$

As a curiosity we might mention here that $i^n \phi_n(-i)$ is the $(n+1)^{\text{th}}$ Fibonacci number, for

$$i \phi_1(-i) = 1, \quad i^2 \phi_2(-i) = 2,$$

and, by expanding the determinant $\phi_n(-i)$ by the first row we obtain,

$$i^n \phi_n(-i) = i^{n-1} \phi_{n-1}(-i) + i^{n-2} \phi_{n-2}(-i).$$

It follows from (1.1) that the n^{th} Fibonacci number may be expressed in the form

$$\prod_{r=1}^{n-1} \left(1 - 2i \cos \frac{r\pi}{n} \right).$$

§ 2. On writing $z = xy + 2$, $a = \frac{1}{2}(x+y)$, it appears that if we write

$$T_n(z, a) \equiv \begin{bmatrix} (z-1), & 2a, & 1 & & & \\ 2a & , & z, & 2a, & 1 & \\ 1 & , & 2a, & z, & 2a, & 1 \\ . & & 1, & 2a, & z, & 2a, & 1 \\ . & . & . & . & . & . & . \\ & & & & & & 1, & 2a, & (z-1) \end{bmatrix}$$

then $T_n(z, a) = P_n(x) P_n(y)$. Accordingly, taking determinants of both sides,

$$\begin{aligned}
 |T_n(z, a)| &= \phi_n(x) \phi_n(y) \\
 &= \prod_{k=1}^n \left(x - 2 \cos \frac{k\pi}{n+1} \right) \prod_{k=1}^n \left(y - 2 \cos \frac{k\pi}{n+1} \right) \\
 &= \prod_{k=1}^n \left(z - 2 - 4a \cos \frac{k\pi}{n+1} + 4 \cos^2 \frac{k\pi}{n+1} \right) \\
 &= \prod_{k=1}^n \left[z - (2 + a^2) + \left(a - 2 \cos \frac{k\pi}{n+1} \right)^2 \right] \quad (2.1)
 \end{aligned}$$

Hence the roots of the equation $|T_n(z, a)| = 0$, regarded as a polynomial of degree n in z are

$$z_k = 2 + a^2 - \left(a - 2 \cos \frac{k\pi}{n+1} \right)^2, \quad k = 1, \dots, n.$$

In particular, when $a = -2$, these roots are

$$z_k = 6 - 16 \cos^2 \frac{k\pi}{2(n+1)}, \quad k = 1, \dots, n.$$

A similar argument applied to the circulant

$$\begin{vmatrix}
 x & 1 & & & 1 \\
 1 & x & 1 & & \\
 & 1 & x & 1 & \\
 . & . & . & . & . \\
 1 & & & 1 & x
 \end{vmatrix} = \prod_{k=1}^n \left(x + 2 \cos \frac{2k\pi}{n} \right),$$

shows that

$$\begin{vmatrix}
 z, & 2a, & 1 & & & 1, & 2a \\
 2a, & z, & 2a, & 1 & & & 1 \\
 1, & 2a, & z, & 2a, & 1 & & \\
 & 1, & 2a, & z, & 2a, & 1 & \\
 . & . & . & . & . & . & . \\
 2a, & 1 & & & & 1, & 2a, & z
 \end{vmatrix} \\
 = \prod_{k=1}^n \left[z - (2 + a^2) + \left(a + 2 \cos \frac{2k\pi}{n} \right)^2 \right] \quad (2.2)$$

Indeed, any circulant can be factorized as follows. The latent roots of

$$\begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \dots & \alpha_{n-1} \\ \alpha_{n-1} & \alpha_0 & \alpha_1 & \dots & \alpha_{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_0 \end{bmatrix} = \alpha_0 I + \alpha_1 K + \dots + \alpha_{n-1} K^{n-1}$$

are $\alpha_0 + \alpha_1 e^{2\pi k i/n} + \dots + \alpha_{n-1} e^{2(n-1)\pi k i/n}$, $k = 1, \dots, n$, since, by a theorem of Frobenius (MacDuffee, 1933, p. 23), those of

$$K \equiv \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad (2.3)$$

are $e^{2\pi k i/n}$, $k = 1, \dots, n$. If the circulant is symmetric, that is, if $\alpha_r = \alpha_{n-r}$, these latent roots are

$$\alpha_0 + \alpha_1 (e^{2\pi k i/n} + e^{-2\pi k i/n}) + \alpha_2 (e^{4\pi k i/n} + e^{-4\pi k i/n}) + \dots,$$

or

$$\alpha_0 + 2\alpha_1 \cos \frac{2k\pi}{n} + 2\alpha_2 \cos \frac{4k\pi}{n} + \dots$$

the last term being $2\alpha_s \cos \frac{2sk\pi}{n}$ if $n = 2s + 1$, or $(-1)^k \alpha_s$ if $n = 2s$. The latent roots are of course factors of the determinant.

The evaluation of the determinant of the matrix

$$Q_n(z, a) \equiv \begin{bmatrix} z, & 2a, & 1 & & & \\ 2a, & z, & 2a, & 1 & & \\ 1, & 2a, & z, & 2a, & 1 & \\ & 1, & 2a, & z, & 2a, & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ & & & & 1, & 2a, & z \end{bmatrix}$$

presents greater difficulties, and so far it has not been found possible to express the determinant as a product of factors linear in z .

Using the addition theorem for determinants, it is easily shown that

$$|T_n| = |Q_n| - 2|Q_{n-1}| + |Q_{n-2}|.$$

Defining $|T_0| \equiv |Q_0| \equiv 1$, $T_1(z, a) \equiv z - 2$, the reciprocal relation is

$$|Q_n| = |T_n| + 2|T_{n-1}| + 3|T_{n-2}| + \dots + (n+1)|T_0|.$$

Now

$$T_n(z, a) = \phi_n(x) \phi_n(y) = \frac{\sin(n+1)\theta \sin(n+1)\psi}{\sin \theta \sin \psi},$$

where $x = 2 \cos \theta$ and $y = 2 \cos \psi$ are the roots of the quadratic equation

$$z = 2 + a^2 - (a - \xi)^2$$

in ξ . We may therefore write, putting $2\alpha = \theta - \psi$, $2\beta = \theta + \psi$,

$$|Q_n| = \sum_{s=1}^{n+1} \frac{(n+2-s) \sin s\theta \sin s\psi}{\sin \theta \sin \psi} = \sum_{s=1}^{n+1} \frac{(n+2-s)(\cos 2s\alpha - \cos 2s\beta)}{\cos 2\alpha - \cos 2\beta}.$$

Some trigonometrical manipulation, which need not be reproduced here, then shows that

$$\begin{aligned} |Q_n| &= \frac{[1 - \cos(n+2)\theta \cos(n+2)\psi] \sin \theta \sin \psi - [1 - \cos \theta \cos \psi] \sin(n+2)\theta \sin(n+2)\psi}{2 \sin \theta \sin \psi (\cos \theta - \cos \psi)^2} \\ &= \frac{1}{2(\cos 2\alpha - \cos 2\beta)} \left[\frac{\sin^2(n+2)\alpha}{\sin^2 \alpha} - \frac{\sin^2(n+2)\beta}{\sin^2 \beta} \right]. \quad (2.4) \end{aligned}$$

The formulae (2.1), (2.2) and (2.4) may be expected to find application in one-dimensional problems in which there are a large number of uniformly spaced "atoms" or nodes, each node being influenced by its two nearest neighbours and also by its two next nearest neighbours. The slight differences between the three determinants considered arise from different end conditions.

§ 3. It is interesting to observe how these determinants may arise in relaxation problems. By Taylor's theorem we find that

$$f(-2a) - 4f(-a) + 6f(0) - 4f(a) + f(2a) = a^4 f^{(iv)}(0) + \frac{1}{2} a^6 f^{(vi)}(0) + \dots$$

Hence if the nodes, at distance a apart, are chosen close enough for a^6 to be neglected, then the relation

$$f(-2a) - 4f(-a) + 6f(0) - 4f(a) + f(2a) = 0$$

is a finite difference approximation at $x=0$ to the differential equation $f^{(iv)}(x) = 0$. In the same way, we can show that another approximation to the same equation is

$$a^2 f''(-a) - 2f(-a) + 5f(0) - 4f(a) + f(2a) = 0.$$

It follows that if $u = f(x)$ and if u and u'' are known on the boundaries $x=0$ and $x=(n+1)a$, then we have the following equations for $u_1 = f(a)$, . . . $u_n = f(na)$,

$$\begin{aligned} 5u_1 - 4u_2 + u_3 &= 2u_0 - a^2 u_0'' \\ -4u_1 + 6u_2 - 4u_3 + u_4 &= -u_0 \\ u_1 - 4u_2 + 6u_3 - 4u_4 + u_5 &= 0 \\ \dots \dots \dots \end{aligned}$$

In the matrix form these equations become

$$T_n(6, -2) \mathbf{u} = \mathbf{v}, \quad (3.1)$$

where \mathbf{u} is the column vector of the unknown elements u_1, \dots, u_n and \mathbf{v} is the column vector with elements

$$2u_0 - a^2 u_0'', -u_0, 0, \dots, 0, -u_{n+1}, 2u_{n+1} - a^2 u_{n+1}''$$

which are determined by the boundary conditions.

Once we have found the inverse (§ 5) of the matrix $T_n(z, a)$, we can write down the solution of the relaxation equations in the form

$$\mathbf{u} = [T_n(6, -2)]^{-1} \mathbf{v} \quad (3.2)$$

These results concerning relaxation equations can be extended, but we shall not dwell on them here.

§ 4. The matrix $T_n(z, a)$ refers to a one-dimensional lattice of "atoms" or nodes in which each node is affected by its nearest and its next nearest nodes. We shall now consider the corresponding matrix for a two-dimensional square lattice. It is a matrix of order mn and is of the form

$$T_m(Z_n, A_n) = P_m[P_n(x)] P_m[P_n(y)] \\ = \begin{bmatrix} Z_n - I_n & 2A_n & I_n & & & \\ & 2A_n & Z_n & 2A_n & I_n & \\ & I_n & 2A_n & Z_n & 2A_n & I_n \\ & & & & & \\ & & & & & \\ & & & & & Z_n - I_n \end{bmatrix}$$

where

$$Z_n = P_n(x) P_n(y) + 2I_n = T_n(z, a) + 2I_n = T_n(z + 2, a),$$

$$A_n = \frac{1}{2}[P_n(x) + P_n(y)] = P_n[\frac{1}{2}(x + y)] = P_n(a)$$

and I_n is the unit matrix of order n .

In fact by equation (4.1) of the author's previous paper,

$$|T_m(Z_n, A_n)| = |P_m(P_n(x))| |P_m(P_n(y))| \\ = \prod_{k=1}^m \prod_{l=1}^n \left(x - 2 \cos \frac{k\pi}{m+1} - 2 \cos \frac{l\pi}{n+1} \right) \left(y - 2 \cos \frac{k\pi}{m+1} - 2 \cos \frac{l\pi}{n+1} \right) \\ = \prod_{k,l} \left[z - (2 + a^2) + \left(a - 2 \cos \frac{k\pi}{m+1} - 2 \cos \frac{l\pi}{n+1} \right)^2 \right].$$

It follows that the roots of the equation $|T_m(Z_n, A_n)| = 0$ are

$$2 + a^2 - \left(a - 2 \cos \frac{k\pi}{m+1} - 2 \cos \frac{l\pi}{n+1} \right)^2, \quad k = 1, \dots, m; \quad l = 1, \dots, n.$$

In particular, if $a = -4$, the roots are

$$18 - 16 \left(\cos^2 \frac{k\pi}{2(m+1)} + \cos^2 \frac{l\pi}{2(n+1)} \right)^2, \quad k = 1, \dots, m; \quad l = 1, \dots, n.$$

A matrix which might be termed a two-dimensional circulant is one of the following type:

$$S = \begin{bmatrix} B_0 & , & B_1 & , & B_2 & , & \dots & , & B_{m-1} \\ B_{m-1} & , & B_0 & , & B_1 & , & \dots & , & B_{m-2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ B_1 & , & B_2 & , & B_3 & , & \dots & , & B_0 \end{bmatrix} = \sum_{r=0}^{m-1} K'_m \langle B_r \rangle,$$

where $X \langle Y \rangle$ denotes a direct product, K_m of order m has the same significance as in (2.3), and where

$$B_j = \alpha_{j0} I_n + \alpha_{j1} K_n + \dots + \alpha_{j, n-1} K_n^{n-1}.$$

Thus,

$$S = \sum_{r=0}^{m-1} \sum_{s=0}^{n-1} \alpha_{rs} K'_m \langle K_n^s \rangle.$$

By a theorem of Stéphanos (1900) the latent roots of S are

$$\sum_{r=0}^{m-1} \sum_{s=0}^{n-1} \alpha_{rs} \lambda_h^r \mu_k^s, \quad h = 1, \dots, m; \quad k = 1, \dots, n.$$

where $\lambda_1, \dots, \lambda_m$ are the latent roots of K_m and μ_1, \dots, μ_n are the latent roots of K_n . Since the latent roots of K_n are $e^{2\pi ki/n}$, $k = 1, \dots, n$, those of S are

$$\sum_{r=0}^{m-1} \sum_{s=0}^{n-1} \alpha_{rs} e^{2\pi i \left\{ \frac{hr}{m} + \frac{ks}{n} \right\}}, \quad h = 1, \dots, m; \quad k = 1, \dots, n.$$

In particular, if $B_0 = Z_n$, $B_1 = B_{m-1} = 2A_n$, $B_2 = B_{m-2} = I_n$ and all other B_r are zero, the roots of the equation $|S| = 0$ are

$$z_{h,k} = 2 + a^2 - \left(a + 2 \cos \frac{2h\pi}{m} + 2 \cos \frac{2k\pi}{n} \right)^2, \\ h = 1, \dots, m; \quad k = 1, \dots, n$$

This S is the appropriate matrix for a closed "square" lattice covering a torus, nearest and next nearest neighbours being considered.

The above formulae can easily be extended to cover three-dimensional types of determinant related to a cubic lattice.

§ 5. In this section we shall consider the inverses of certain matrices. A knowledge of these may be valuable for two reasons. First, they may be used to solve a system of relaxation equations such as (3.1). Secondly,

if one of these inverse matrices arises in some physical problem we can at once obtain its latent roots if those of the original matrix are known.

Let $R_n(x, a, b)$ be the matrix

$$\begin{bmatrix} x+b & 1 & & & & \\ & 1 & x & 1 & & \\ & & 1 & x & 1 & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & & & 1 & x & 1 \\ & & & & & 1 & x+a \end{bmatrix}$$

of order n , and let us extend our definition of $\phi_n(x)$ so that, even when n is not integer,

$$\phi_n(2 \cos \theta) = \frac{\sin (n+1)\theta}{\sin \theta}, \quad \theta \neq 0, \pi$$

$$\phi_n(2) = n+1, \quad \phi_n(-2) = (-1)^n(n+1)$$

It can then be shown that the r, s^{th} element η_{rs} of the inverse of $R_n(x, -\phi_{\beta-1}(x)/\phi_{\beta}(x), -\phi_{\alpha-1}(x)/\phi_{\alpha}(x))$ is given by

$$\eta_{rs} = \eta_{sr} = \frac{(-1)^{r+s} \phi_{\alpha+r-1}(x) \phi_{\beta+n-s}(x)}{\phi_{\alpha+\beta+n}}, \quad r \leq s \quad (5.1)$$

In illustration, we observe that if $x = -2$, $\alpha = -\frac{1}{2}$, $\beta \rightarrow \infty$, then

$$-2[R_n(-2, +1, -1)]^{-1} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 3 & 3 & \dots & 3 \\ 1 & 3 & 5 & \dots & 5 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 3 & 5 & \dots & 2n-1 \end{bmatrix}$$

Since the latent roots of $R_n(-2, +1, -1)$ are $-4 \cos^2[(2r-1)\pi/4n]$, those of the matrix on the right-hand side of the last equation are $\frac{1}{2} \sec^2 [(2r-1)\pi/4n]$, $r = 1, \dots, n$. This matrix and its latent roots arise in certain calculations by Duncan (1950).

Again, putting $\alpha = \beta = 0$, we obtain the formula

$$\eta_{rs} = \eta_{sr} = \frac{(-1)^{r+s} \phi_{r-1}(x) \phi_{n-s}(x)}{\phi_n(x)}, \quad r \leq s$$

for the inverse of $P_n(x)$, or $R_n(x, 0, 0)$. In particular, the inverse matrix of $P_n(-2)$ is given by

$$\eta_{rs} = \eta_{sr} = \frac{-r(n+1-s)}{n+1}, \quad r \leq s$$

This last formula was given, apart from a misprinted sign, by Todd (1950).

Since $P_n(x)$ and $P_n(y)$ are commuting matrices, and since $T_n(z, a) = P_n(x) P_n(y)$, where x and y are the roots of the quadratic equation

$$(\xi - a)^2 + z - 2 - a^2 = 0$$

in ξ , we have at once,

$$[T_n(z, a)]^{-1} = [P_n(x)]^{-1} [P_n(y)]^{-1}.$$

For many purposes it is convenient to leave T_n^{-1} in this form. Thus (3.2) may be written

$$u = [P_n(-2)]^{-1} [P_n(-2)]^{-1} v,$$

from which the components of u are easily obtained, by premultiplying the known vector v by $[P_n(-2)]^{-1}$ twice in succession.

It is difficult to find the inverse matrix of $P_m[P_n(x)]$ in a form suitable for calculation when m and n are large, but the following method can be adopted when m and n are small.

Let $Q_m(x)$ be the adjugate of $P_m(x)$. Then

$$P_m(x) Q_m(x) = \phi_m(x) I_m.$$

Consequently

$$\begin{aligned} P_m(P_n(x)) Q_m(P_n(x)) &= I_m \langle \phi_m(P_n(x)) \rangle \\ &= I_m \left\langle \prod_{k=1}^m \left[P_n(x) - 2 \cos \frac{k\pi}{m+1} I_n \right] \right\rangle \\ &= I_m \left\langle \prod_{k=1}^m P_n \left(x - 2 \cos \frac{k\pi}{m+1} \right) \right\rangle. \end{aligned}$$

It follows that

$$\begin{aligned} [P_m(P_n(x))]^{-1} &= Q_m(P_n(x)). \quad I_m \left\langle \prod_{k=1}^m \left[P_n \left(x - 2 \cos \frac{k\pi}{m+1} \right) \right]^{-1} \right\rangle \\ &= Q_m(P_n(x)). \quad I_m \left\langle \prod_{k=1}^m \frac{Q_n \left(x - 2 \cos \frac{k\pi}{m+1} \right)}{\phi_n \left(x - 2 \cos \frac{k\pi}{m+1} \right)} \right\rangle; \end{aligned}$$

the matrix $Q_n(x)$ is of course obtainable from the formula (5.1).

§ 6. In conclusion, it is interesting to observe that we can write

$$P_n(x) = xI_n + 2 \cos \Theta_n,$$

where

$$\Theta_n = A_n \Psi_n A_n,$$

in which

$$\Psi_n = \frac{\pi}{n+1} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n \end{bmatrix}$$

and $A_n (= A_n^{-1})$ is the matrix whose r, s^{th} element is $\sqrt{\left(\frac{2}{n+1}\right)} \sin \frac{rs\pi}{n+1}$.
To show this it is only necessary to verify that

$$A_n P_n(0) A_n = 2 \cos \Psi$$

and that

$$A_n^2 = I_n.$$

It also follows easily from these formulae that the latent vectors of $P_n(x)$ are the columns of A_n .

It can be shown that the elements θ_{rs} of Θ_n are given by

$$\theta_{rs} = 0, \text{ if } r+s \text{ is even and } r \neq s, \theta_{rr} = \frac{\pi}{2},$$

$$\theta_{rs} = \frac{\pi}{2(n+1)^2} \left(\cot^2 \frac{(r+s)\pi}{(n+1)2} - \cot^2 \frac{(r-s)\pi}{(n+1)2} \right), \text{ if } r+s \text{ is odd.}$$

REFERENCES TO LITERATURE

- DUNCAN, W. J., 1950. "Dependence of Errors in the Natural Frequencies of the Numbers of Elements in a Segmental Representation of a Uniform Shaft". *College of Aeronautics, Cranfield. Report.*
- MACDUFFEE, C. C., 1933. *The Theory of Matrices*, Berlin.
- RUTHERFORD, D. E., 1947. "Some Continuant Determinants Arising in Physics and Chemistry", *Proc. Roy. Soc. Edin.*, A, LXII, 229-230.
- STEPHANOS, C., 1900. "Sur une extension du calcul des substitutions linéaires", *Journ. de Mathématiques*, V, VI, 73-128.
- TODD, JOHN, 1950. "The Condition of a Certain Matrix", *Proc. Camb. Phil. Soc.*, XLVI, 116-118.

(Issued separately February 9, 1952)

XVII.—The Sargent Diagram for the Electron-capture Process, and the Disintegration Energies of Heavy β -emitters.*

By N. Feather, Department of Natural Philosophy, University of Edinburgh

(MS. received May 23, 1951. Read November 5, 1951)

SYNOPSIS

A Sargent diagram is presented containing 12 plotted points relative to capture-active species in the range of atomic number (Z) from 80 to 98 inclusive. Arguments are adduced to show that the "allowed" line of the diagram is located as theory predicts, and the capture transformations of other heavy capture-active species are discussed with the aid of the diagram. In particular, values are deduced for the energies of capture transformation of 17 species for which $79 \leq Z \leq 85$, and, taking count of these values, the energies of β -disintegration of 150 species having $76 \leq Z \leq 98$ are assumed known, and are suitably plotted against neutron number N . Discontinuities are found, for certain values of isotopic number, in the region of $N = 126$ (and $Z = 82$). Values of α -disintegration energy are also deduced for certain isotopes of bismuth and lead.

THE SARGENT DIAGRAM

THE original Sargent diagram (Sargent, 1933) provided the first successful demonstration of the existence of a significant relation between disintegration energy and disintegration constant in β -emission. At a time when there was no satisfactory theory of the disintegration process, when the distinction between simple and complex β -spectra was at the best very imperfectly appreciated, and when the only β -active bodies known were those of classical radioactivity, this essentially successful essay in empiricism was based upon 12 plotted points—of which one (for $^{211}_{82}\text{AcB}$) appeared obviously out of place. With the other 11 points it exhibited the distinction between "allowed" and "forbidden" transitions, which has since been amply vindicated, and it spoke in favour of the diagram that further investigation at an early stage (Sargent, 1939) showed that the apparently anomalous point was incorrectly placed as a result of faulty interpretation of experimental data. From the vantage-point of to-day we know that the surprising success of the diagram was to some extent fortuitous; it derived from the fact that in several cases the β -active bodies concerned transform predominantly according to a single mode, and from the now theoretically justified result that in β -disintegration (as opposed to α -disintegration) transition probabilities

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

are not very strongly dependent on nuclear charge. Only in this way could significant regularities have emerged in a diagram covering the range in atomic number, Z , $80 < Z < 92$. From the vantage-point of to-day, also, we know that the multiplicity of transition types is greater than the two of the original diagram (Feather and Richardson, 1948), but its essential features are still recognizable in its modern counterpart, and its influence on the development of theory cannot be overestimated.

In relation to the electron-capture process the empiricist is still to-day very much in the position in which Sargent found himself in respect of β -disintegration in 1933. Because of the nature of the case capture-disintegration energies cannot be determined directly, but have to be deduced from indirect evidence, and, again, that evidence is at present available in any quantity only for the heavy radioelements. To complicate the position there is the prediction of theory (Konopinski, 1943) that the Z -dependence of disintegration probability is more marked for capture-transformation than for β -disintegration. In spite of all this, the attempt to construct a Sargent diagram for the capture process would appear to be worth while. Such an attempt is made here, and one of the uses of the diagram so obtained is illustrated in the second part of the paper.

To the writer's knowledge, the only previous attempt in the direction indicated is that of Thompson. In a short summary of his work Thompson (1949) gives a diagram containing 15 plotted points relative to heavy capture-active species (which unfortunately are not identified in the published account), and suggests that 11 of these points might define a single Sargent curve. The other 4 points appear to represent transitions which, in comparison with the former, are more forbidden. Transformation energies are calculated by the method of "closed cycles" using known (or inferred) values of α -disintegration energy, and in each case the point plotted represents the total energy available for K -capture transformation from nuclear ground state to ground state. The author recognizes the possibility that in many cases the actual transformation energy may be much less than the assumed energy, but in the absence of information concerning associated γ -ray emission he takes the bold step of disregarding this possibility. The slope of his main Sargent line corresponds to a fourth-power law relating disintegration constant with disintegration energy.

Now it is a definite prediction of theory that the transition probability for an allowed capture transformation is directly proportional to the square of the energy of the emitted neutrino; thus if Thompson's line is significant, either theory is incorrect on this point or no allowed transitions are represented on his diagram. It would appear more likely than either alternative that the line on his diagram is not a significant Sargent line, and that the unjustified assumption regarding disintegration energy

above referred to has resulted in the serious misplacing of several points on the diagram. For that reason the information available concerning the heavy capture-active species has been re-surveyed quite independently of Thompson's earlier work, and to the conclusions reached notes are appended concerning each case treated, so that the reader may judge for himself how much is still conjecture and how much established fact. During the last two years much has been discovered concerning the γ -radiation emitted by capture-active species, but conjecture has still to play a part in any survey, as will be evident from what follows. Even admitting some conjecture, however, only 12 points appear on the Sargent diagram here proposed.

As in Thompson's survey, capture-disintegration energies are again deduced by the standard method of the closed decay cycle. In this method, for each series of capture-active species of constant isotopic number, one energy of β -disintegration and all relevant energies of α -disintegration must be known. These conditions are satisfied for some 26 capture-active species for which $89 \leq Z \leq 98$ —and, with confidence, for none having $Z < 87$. The present diagram then refers only to $89 \leq Z \leq 98$. Closing of a decay cycle effectively gives the total energy E available for capture of a "valency" electron by the nucleus. For K -electron capture, if W_K is the K -ionization energy of the daughter atom, then $E - W_K$ is energy carried away by the neutrino in a ground-to-ground state transition, and $E - W_K - E'_r$ is the energy so carried away when the daughter nucleus is left excited in a state of excitation energy E'_r . Table I shows the values of E , W_K and E'_r (in keV) used in plotting the points on the diagram. As previously mentioned, only

TABLE I

Species	E	W_K	E'_r	Reference, fig. 1
$^{231}_{91}U$	520 ± 20	113	0	1
$^{231}_{93}Np$	430 ± 15	116	0	2
$^{231}_{95}Am$	1,060	122	0	3
	1,060	122	277	4
$^{227}_{89}Ac$	1,350	104	0	5
$^{227}_{91}Pa$	510	110	0	6
$^{235}_{92}U$	540 ± 20	113	0	7
$^{235}_{94}Pu$	180 ± 100	119	0	8
$^{237}_{94}Pu$	170 ± 100	119	0	9
$^{230}_{91}Pa$	$1,420 \pm 200$	110	940	10
$^{234}_{93}Np$	$2,240 \pm 300$	116	1,600	11
$^{234}_{95}Of$	300 ± 200	133	0	12

12 points are plotted, these having reference to the assumed chief modes of capture transformation of 11 of the 26 species with $89 \leq Z \leq 98$ for which E has been deduced. For the remaining 15 species in this group information regarding disintegration modes was judged too indefinite to enable a plot to be made.

In the table itself horizontal lines divide the data into three groups according to estimated reliability—the estimate taking count not only of the reliability of energy determinations, but also of the degree of confidence in the assignment of partial disintegration constants as discussed in the notes below. In the Sargent diagram the four “more reliable” points are represented by full circles, the five “reliable” points by full triangles and the three “less reliable” points by open circles. Notes on individual cases are as follows, τ being the half-value period directly determined for each species, and the figure in brackets referring to the representative point on fig. 1.

- ${}^{231}_{91}\text{U}$. $\tau = 4.2d$. (1) α /capture branching ratio $\sim 5 \times 10^{-5}$. Capture assumed to occur in a single mode, in spite of knowledge of 35 keV state in ${}^{231}_{91}\text{Pa}$ from the disintegration ${}^{231}_{90}\text{U} \rightarrow {}^{231}_{91}\text{Pa}$ (Knight and Macklin, 1949). “No γ -rays” reported by Crane, Ghiorso and Perlman (1949).
- ${}^{234}_{93}\text{Np}$. $\tau = 420 \pm 20d$. (2) α /capture branching ratio $\sim 5 \times 10^{-5}$. Capture assumed to occur in a single mode. Energy almost certainly insufficient to excite the 410 keV state of ${}^{234}_{93}\text{U}$ (Sullivan, Kohman and Swartout, 1945) by K -capture. Excitation of the postulated 50 keV state (Albony and Teillac, 1951) is not, however, ruled out.
- ${}^{241}_{95}\text{Am}$. $\tau = 12h$. (3 and 4) α /capture branching ratio $\sim 10^{-4}$. Seaborg, James and Morgan (1948) report a 285 keV γ -ray in about 60 per cent. of disintegrations. This is tentatively identified with the 277 keV γ -ray emitted in the disintegration ${}^{241}_{95}\text{Np} \rightarrow {}^{241}_{96}\text{Pu}$ (Fulbright, 1947), and capture therefore assumed to be 60 per cent. to the 277 keV state and 40 per cent. to the ground state of ${}^{241}_{96}\text{Pu}$. This assumption is roughly consistent with knowledge of the relative intensities of the 400 keV and 680 keV components of the β -spectrum of ${}^{241}_{95}\text{Np}$.
- ${}^{244}_{90}\text{Ac}$. $\tau = 2.9h$. (5) α /capture branching ratio $\sim 1/9$. Capture assumed in roughly equal proportions to the ground state and to the state (or states) at ~ 85 keV; compare the α -disintegration ${}^{244}_{90}\text{RdTh} \rightarrow {}^{244}_{91}\text{ThX}$. Only one point (for the ground-to-ground state transition) is plotted on the diagram.
- ${}^{248}_{94}\text{Pa}$. $\tau = 1.5d$. (6) α /capture branching ratio $\sim 1/99$. Capture is assumed in 60/40 ratio between transitions to the ground state and to the 80 keV state (Studier, 1947) of ${}^{248}_{95}\text{Th}$. Point for former transition alone is plotted. Possible excitation of the 310 keV state of ${}^{248}_{95}\text{Th}$ is neglected.
- ${}^{252}_{98}\text{U}$. $\tau = 9.3 \pm 0.5m$. (7) α /capture branching ratio $\sim 4/1$. Nothing is known about the excited states of ${}^{252}_{99}\text{Pa}$, but on account of the smallness of the disintegration energy it is provisionally assumed that capture is by a single mode to the ground state of that nucleus.
- ${}^{256}_{98}\text{Pu}$. $\tau = 9.0 \pm 0.5h$. (8) α /capture branching ratio $\sim 1/30$. Nothing is known about the excited states of ${}^{256}_{99}\text{Np}$, but the disintegration energy

is again very small, and capture by a single mode to the ground state is once more postulated.

- $^{234}_{94}\text{Pu}$. $\tau = 40d$. (9) α /capture branching ratio very small. James, Thompson and Hopkins (1948) report "no γ -rays". Disintegration energy is very small and capture is therefore assumed to be by a single mode to the ground state. The possibility of excitation of the 60 keV state of $^{234}_{94}\text{Np}$ (Melander and Slätis, 1948; Seaborg, James and Morgan, 1948) is neglected.
- $^{230}_{90}\text{Pu}$. $\tau = 17d$. (10) β /capture branching ratio $\sim 1/9$, α /capture branching ratio $\sim 3 \times 10^{-6}$. Osborne, Thompson and Van Winkle (1946) report γ -ray of 940 keV energy. Capture assumed, very tentatively, to be to the 940 keV state of $^{230}_{90}\text{Po}$ in about 70 per cent. of disintegrations.
- $^{234}_{93}\text{Np}$. $\tau = 4.5d$. (11) α /capture branching ratio very small. James, Florin, Hopkins and Ghiorso (1948) report a γ -ray of 1.8 MeV energy. This is tentatively identified with the 1.5 MeV γ -ray emitted in the disintegration $^{234}_{93}\text{U} \rightarrow ^{234}_{92}\text{U}$ (Bradt and Scherrer, 1945, 1946), and a large fraction of the capture disintegrations is assumed to be to a 1.6 MeV state of $^{234}_{93}\text{U}$ in this case.
- $^{244}_{98}\text{Cf}$. $\tau \sim 45m$. (12) α /capture branching ratio probably somewhat greater than 1. Nothing is known about the excited states of $^{244}_{98}\text{Bk}$, but the smallness of the disintegration energy makes it likely that capture is by a single mode to the ground state.

The Sargent diagram for the 11 species treated in the notes is given in fig. 1. Obviously the point for $^{238}_{92}\text{U}$ is the most well-attested point for a transition which can provisionally be regarded as "allowed". On that assumption the straight line 0 is drawn with the theoretical slope of 2 through this point. It will be seen that within the indicated uncertainty of present experimental knowledge the points for $^{244}_{98}\text{Cf}$ and $^{234}_{94}\text{Pu}$ also lie on this line, and that no points lie significantly above it. Further, the three species $^{244}_{98}\text{Cf}$, $^{234}_{94}\text{Pu}$ and $^{238}_{92}\text{U}$ are the only even-even species represented on the diagram. It is characteristic of the Sargent diagram for the heavy β -active species, also, that the points for even- Z bodies in general are concentrated along the upper curves of the diagram. In fig. 1, when line 0 has been drawn, 8 of the remaining 9 points can be accounted for, almost within the assigned limits of experimental uncertainty, by the straight line II. The slope of this line is considerably greater than 2—as befits the line for a forbidden transition—and it has been labelled II rather than I on the assumption that the first-forbidden line (as in the Sargent diagram for heavy β -emitters) is in fact parallel to the allowed line and not very far removed from it, and the second-forbidden line is the first to show a greater slope (and, in a definitive diagram, to be curved rather than straight)*. Beyond line II on fig. 1

* The terms "first-forbidden", "second-forbidden", etc., are used throughout this paper entirely empirically. They correspond to a classification more directly in relation to spin change (by 2 units for second-, by 3 units for third-forbidden transitions, etc.) than to parity change.

the point for $^{235}_{93}\text{Np}$ is well attested, and is very probably indicative of the position of the third-forbidden line shown tentatively as III. It may be noted here that the points for $^{230}_{91}\text{Pa}$ and $^{234}_{93}\text{Np}$ would both lie nearer to line III (and in fact rather to the right of it) than to line II, if ground-to-ground state disintegrations were postulated as the predominant modes in these two cases. Only further study of the intensities of the γ -radiations concerned can settle these ambiguities.

As regards the absolute position of the allowed line 0 on fig. 1, the approximate method of calculating values of the "comparative half-life", ft , of the Fermi theory of β -disintegration recently given by Moszkowski (1951) has been used to compare the degrees of forbiddenness represented by this line and by the allowed line on the diagram for β -emitters having $81 \leq Z \leq 83$ given by Feather and Richardson (1948).

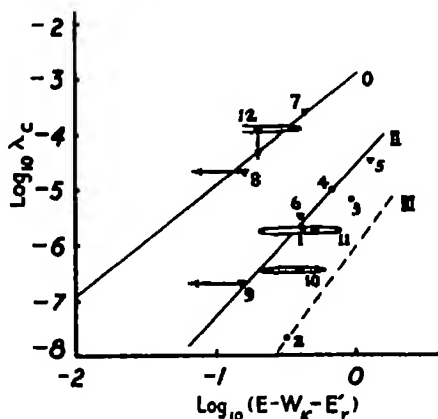


FIG. 1.—Sargent diagram for capture-active species having $80 < Z < 98$.
 λ_c : partial disintegration constant for K -electron capture (in sec.^{-1}).
 $E - W_K - E'_\nu$: neutrino (transition) energy (in MeV).

Taking 93 as the effective value of Z for the purpose of fig. 1 of this paper, and $Z = 82$ for the Sargent diagram previously published, we obtain $\log_{10}(ft) = 4.6$ for the line representing the capture disintegrations which we now assume to be allowed disintegrations, and $\log_{10}(ft) = 5.2$ for that representing those β -disintegrations generally accepted as correctly so classified. The agreement between these two values is very clearly encouraging*. We can reasonably conclude that the allowed line has been fixed with fair precision on fig. 1.

Further support for the correctness of our identification is obtained from a consideration of the capture transformation of the 15 species having $80 \leq Z \leq 98$ for which evidence regarding disintegration modes

* It is possibly significant that the former value is less than the latter, since the partial disintegration constants here used in fact include contributions from L -, M -electron capture, etc.

was considered too indefinite for use in the construction of fig. 1. For some of these species capture no doubt proceeds predominantly by ground-to-ground state transitions. If we make this assumption for each in turn we effectively place the representative point on fig. 1 as high on the diagram, and as far to the right, as it can possibly be placed. In this way we find that no new point is placed to the left of the line 0, and that three points fall sufficiently near to this line (though to the right of it) to suggest that the species in question transform, in a major mode in each case, by allowed (or first forbidden) transitions to the ground state, or to a low-lying excited state, of the product nucleus concerned. The three species for which this behaviour is indicated are $^{225}_{90}\text{Th}$, $^{232}_{93}\text{Np}$ and $^{232}_{94}\text{Pu}$. Only in one respect do these assignments run counter to reported fact: it would appear impossible, if they are accepted, that there should be any prominent component of "hard" γ -radiation associated with the capture transformation of $^{232}_{93}\text{Np}$, as has been suggested (Magnusson, Thompson and Seaborg, 1950).

Concerning the capture transformations of the remaining 12 species in this group nothing very definite can be said so long as our simplifying assumption is retained. Probably more by coincidence than otherwise, the representative points for the capture transformations of $^{223}_{89}\text{Ac}$ and $^{227}_{91}\text{Pa}$ then fall on or very near to line II and those for $^{228}_{91}\text{Pa}$, $^{240}_{95}\text{Am}$ and $^{241}_{96}\text{Cm}$ fall on line III, but when more is known of partial disintegration modes in these cases it may well appear that the ground-to-ground state transitions are more forbidden than these placings would indicate, and that the main disintegration mode in each case is to an excited state of the product nucleus and is less forbidden than here supposed. Our conjectured placing is most likely to prove correct for $^{223}_{89}\text{Ac}$, since the energy available for excitation of the product nucleus is small, and probably least likely to be significant for $^{228}_{91}\text{Pa}$ (disintegration energy 1.92 MeV), because it is known from a study of the β - and γ -radiations of $^{228}_{90}\text{Th}$, that many low-lying states of $^{228}_{90}\text{RdTh}$ are available for excitation by the capture process. With $^{238}_{95}\text{Am}$, capture transformation predominantly to the 1.2 MeV state of $^{238}_{94}\text{Pu}$ revealed in the β -disintegration of $^{238}_{93}\text{Np}$ (Jaffey and Magnusson, 1949) would be represented by a point falling closely on line II, and a point falling on the same line would represent the capture transformation of $^{240}_{95}\text{Am}$, if, as has been suggested (Seaborg, James and Morgan, 1948), the predominant mode in this case results in excitation of a state of 1.3 MeV energy in $^{240}_{94}\text{Pu}$. Possibly these two transformations are those which occur in fact, and are correctly described as second forbidden.

THE DISINTEGRATION ENERGIES OF HEAVY β -EMITTERS

In any survey of nuclear stability, one profitable method of exhibiting stability limits is to plot β -disintegration energy, E_β , against A (mass

number), N (neutron number) or Z , and to draw in the various smooth curves for $A - 2Z$ (isotopic number) constant. This is the method employed by Saha and Saha (1946). Suess (1951) has recently pointed to marked discontinuities in the curves of E_β against N for $A - 2Z = 13$ and $A - 2Z = 25$, occurring at $N = 50$ and $N = 82$, respectively, and has suggested that a similar discontinuity in the appropriate curve probably occurs at $N = 126$. Hitherto data for the construction of the curves of E_β against N in just this region have been lacking through want of knowledge of the energies of capture transformation of certain "key" species. Fig. 1 enables us to make predictions regarding these energies, and the present section will review these predictions, and discuss the curves exhibiting E_β as a function of N which may be constructed on the basis of the values so obtained. The capture-active species in question are $^{198}_{79}\text{Au}$, $^{201}_{81}\text{Tl}$, $^{202}_{81}\text{Tl}$, $^{204}_{81}\text{Tl}$, $^{202}_{82}\text{Pb}$, $^{203}_{82}\text{Pb}$, $^{205}_{82}\text{Pb}$, $^{204}_{83}\text{Bi}$, $^{205}_{83}\text{Bi}$, $^{206}_{83}\text{Bi}$, $^{206}_{83}\text{Bi}$, $^{206}_{84}\text{Po}$, $^{207}_{84}\text{Po}$, $^{208}_{84}\text{Po}$, $^{209}_{85}\text{At}$ and $^{211}_{85}\text{At}$. Within the accuracy required, fig. 1 may be regarded as applicable over the extended range of Z values, $79 \leq Z \leq 98$, although account should strictly be taken of a general shift of the lines of the figure downwards, as Z decreases. For the allowed line 0 the theoretical downwards shift, for a decrease in Z from 93 to 83 is of the order of 0.15 in $\log_{10} \lambda_c$.

Table II sets out the estimated values of E for the capture-active species in the above list (these are the values of the negative energy of β -disintegration for the respective daughter products) and the notes below discuss individual cases as before. One general principle may be stated at the outset. Two of the species ($^{198}_{79}\text{Au}$ and $^{204}_{81}\text{Tl}$) are of even A and odd Z and are both β - and capture-unstable. Since in each case both modes of disintegration give rise to daughter products which are of even A and even Z , it is assumed, for each species, that the ground-to-ground state transitions (β - and capture transitions) are of the same degree of forbiddenness.

TABLE II

Species	$^{198}_{79}\text{Au}$	$^{201}_{81}\text{Tl}$	$^{202}_{81}\text{Tl}$	$^{204}_{81}\text{Tl}$	$^{202}_{82}\text{Pb}$	$^{203}_{82}\text{Pb}$	
$E(\text{MeV})$	< 0.36	0.76	1.47	< 0.18	0.4 ± 0.3	0.61	
Species	$^{205}_{83}\text{Pb}$	$^{205}_{83}\text{Bi}$	$^{206}_{83}\text{Bi}$	$^{206}_{83}\text{Bi}$	$^{207}_{83}\text{Bi}$	$^{208}_{83}\text{Bi}$	$^{208}_{84}\text{Po}$
$E(\text{MeV})$	0.25 ± 0.2	2.7 ± 0.2	1.95	1.84	2.59	< 2.70	0.9
Species	$^{207}_{84}\text{Po}$	$^{208}_{84}\text{Po}$	$^{209}_{85}\text{At}$	$^{211}_{85}\text{At}$			
$E(\text{MeV})$	1.56 ± 0.1	< 0.2	3.1 ± 0.3	0.96			

Notes on individual species are as follows (here E' is written for $E - E'$):

$^{198}_{79}\text{Au}$. $\tau = 2.69d$. β /capture branching ratio > 250 (Renard, 1949). Since the ground-to-ground state β -disintegration is probably third-

forbidden, the same is assumed of the ground-to-ground state capture transformation (*v. sup.*).

- ${}^{201}_{81}\text{Tl}$. $\tau = 72h$. It is assumed that the predominant mode involves a second-forbidden capture transition with $E' - W_K = 0.47 \text{ MeV}$ to an excited state of ${}^{201}_{80}\text{Hg}$ having $E'_r = 0.21 \text{ MeV}$ (Neumann and Perlman, 1950).
- ${}^{202}_{81}\text{Tl}$. $\tau = 13d$. It is assumed that the predominant mode involves a third-forbidden capture transition with $E' - W_K = 0.99 \text{ MeV}$ to an excited state of ${}^{202}_{80}\text{Hg}$ having $E'_r = 0.40 \text{ MeV}$ (Maurer and Ramm, 1942).
- ${}^{204}_{81}\text{Tl}$. $\tau = 2.7y$. $\beta/\text{capture branching ratio} > 50$ (estimated from intensity of soft quantum radiation as measured by Evans, 1950). It is assumed that the ground-to-ground state capture transformation is third-forbidden as the ground-to-ground state β -disintegration appears to be (*v. sup.*).
- ${}^{208}_{82}\text{Pb}$. τ assumed long. Degree of forbiddenness assumed equal to or less than three.
- ${}^{209}_{82}\text{Pb}$. $\tau = 52h$. It is assumed that excited states of ${}^{209}_{81}\text{Tl}$ having $E'_r = 0.47, 0.28 \text{ MeV}$ are both excited in this transformation (Lutz, Pool and Kurbatov, 1944; Slätis and Siegbahn, 1949). Transition to the 0.47 MeV state is assumed allowed with $E' - W_K = 55 \text{ keV}$.
- ${}^{210}_{82}\text{Pb}$. τ assumed very long. Degree of forbiddenness assumed equal to or less than three.
- ${}^{210}_{83}\text{Bi}$. $\tau = 12h$. It is assumed that the predominant mode involves a third-forbidden capture transition with $E' - W_K = 2.4 \pm 0.2 \text{ MeV}$ to an excited state of ${}^{210}_{82}\text{Pb}$ having $E'_r = 0.217 \text{ MeV}$ (Suryar, *et al.*, 1950).
- ${}^{211}_{83}\text{Bi}$. $\tau = 14.5d$. It is assumed that the predominant mode involves an allowed capture transition with $E' - W_K = 25 \text{ keV}$ to an excited state of ${}^{211}_{82}\text{Pb}$ having $E'_r = 1.84 \text{ MeV}$ (Karraker and Templeton, 1950).
- ${}^{212}_{83}\text{Bi}$. $\tau = 6.4d$. It is assumed that the predominant mode involves an allowed capture transition with $E' - W_K = 36 \text{ keV}$ to an excited state of ${}^{212}_{82}\text{Pb}$ having $E'_r = 1.72 \text{ MeV}$ (Alburger and Friedlander, 1951).
- ${}^{214}_{83}\text{Bi}$. $\tau \sim 60y$. The value of E in this case is very insensitive to the degree of forbiddenness assumed for the predominant capture mode, so long as this is to an excited state of ${}^{214}_{82}\text{Pb}$ having $E'_r = 2.49 \text{ MeV}$ (Neumann and Perlman, 1951). For sake of definiteness a value $E' - W_K = 12 \text{ keV}$ (second-forbidden transition) is here assumed.
- ${}^{215}_{83}\text{Bi}$. τ probably very long. It is assumed that the disintegration energy available is insufficient for the excitation of the 2.62 MeV state of ${}^{215}_{82}\text{Pb}$ by a K -capture transformation (Neumann and Perlman, 1951).
- ${}^{216}_{84}\text{Po}$. $\tau = 9d$. $\alpha/\text{capture branching ratio} \sim 1/9$ (Howland, Templeton and Perlman, 1947). It is assumed that the predominant mode involves an allowed capture transition with $E' - W_K = 27 \text{ keV}$ to an excited state of ${}^{216}_{83}\text{Bi}$ having $E'_r = 0.8 \text{ MeV}$ (Templeton, Howland and Perlman, 1947).
- ${}^{217}_{84}\text{Po}$. $\tau = 5.7h$. It is assumed that the predominant mode involves an allowed transition with $E' - W_K = 0.17 \text{ MeV}$ to an excited state of ${}^{217}_{83}\text{Bi}$ having $E'_r = 1.3 \text{ MeV}$ (Templeton, Howland and Perlman, 1947).

- $^{208}_{84}\text{Po}$. $\tau = 2.93\text{y}$. $\alpha/\text{capture branching ratio} > 50$ (Templeton, 1950). It is assumed that the ground-to-ground state capture transformation is of order of forbiddenness not greater than three—unless the $\alpha/\text{capture branching ratio}$ is in fact much greater than the limit given.
- $^{209}_{85}\text{At}$. $\tau = 5.5\text{h}$. $\alpha/\text{capture branching ratio} \sim 1/20$ (Barton, Ghiorso and Perlman, 1951). It is assumed that the predominant mode involves a third-forbidden transition with $E - W_K = 3.0 \pm 0.3 \text{ MeV}$ to the ground state of $^{208}_{84}\text{Po}$.
- $^{211}_{85}\text{At}$. $\tau = 7.5\text{h}$. $\alpha/\text{capture branching ratio } 40.9/59.1$ (Neumann and Perlman, 1951). It is assumed that the predominant mode involves a second-forbidden capture transition with $E - W_K = 0.87 \text{ MeV}$ to the ground state of $^{210}_{84}\text{Po}$.

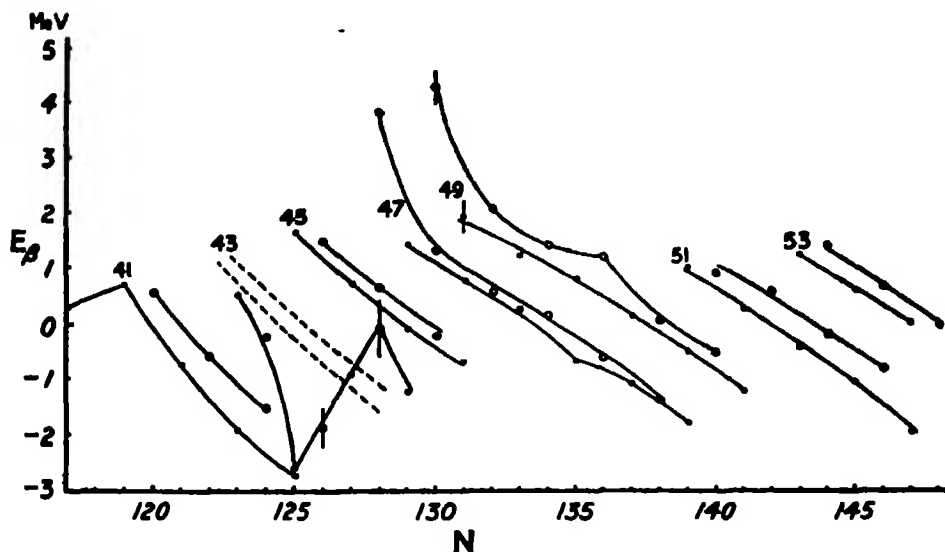


FIG. 2.— β -disintegration energy, E_β , as a function of neutron number, N , for species of odd A having isotopic numbers from 41 to 53 inclusive. Species of odd Z represented by open circles, those of even Z by full circles.

The amount of conjecture still remaining in our conclusions has already been alluded to, and it will be evident from the above notes that it is not small, but that it is not entirely random conjecture will be appreciated from the fact that it has been subject to the following checks. First, in certain cases the method of the closed decay cycle provides a natural check. Thus from a knowledge of α -disintegration energies we have (in MeV)

$$E(^{211}_{85}\text{At}) = E(^{207}_{83}\text{Bi}) - 1.58 \pm 0.02$$

$$E(^{209}_{85}\text{At}) = E(^{205}_{83}\text{Bi}) + 0.80 \pm 0.03,$$

whereas from Table II we have -1.63 and $+1.15 \pm 0.3$ for the numerical terms in these two equalities. Any choice of possibilities, consistent

with the validity of our Sargent diagram, other than the choice indicated in the notes, would not have resulted in such good agreement.

The second check on conjecture is that when resulting values of $-E$ are used in constructing curves of E_β against N for $A-2Z$ constant, these curves shall be as smooth as possible. We pass therefore, directly, to a consideration of these curves, always remembering that this criterion has been adopted in deriving some of the data used in their construction. The curves for odd values of the isotopic number $A-2Z$ from 41 to 53 inclusive are given in fig. 2, those for even isotopic number from 40 to 52 in fig. 3 or fig. 4 depending upon whether the charge numbers of the β -active species involved are odd or even. In terms of proton-neutron

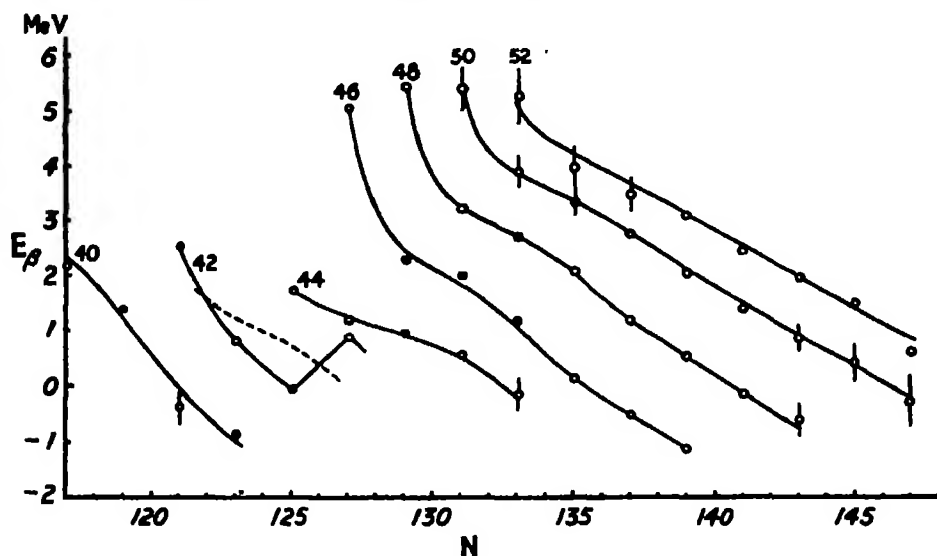


FIG. 3.— E_β as a function of N for species of even A and odd Z . Isotopic numbers from 40 to 52 inclusive.

parities, fig. 2 refers to the β -disintegrations of odd-even and even-odd nuclei, fig. 3 to the β -disintegrations of odd-odd nuclei, and fig. 4 to those of even-even nuclei. It will be realized that some of the values plotted in these figures depend upon measurements of β - and γ -ray energies directly, and others to a greater or less degree upon the use of the method of closed decay cycles with directly measured, or with inferred, values of α -disintegration energy. Where the estimated uncertainty of a plotted point is greater than ± 0.2 MeV the assumed magnitude of the probable error is represented graphically in the usual way. In a few cases total disintegration energies have had to be assumed without fully convincing evidence from β - γ correlation studies or otherwise—and one α -disintegration energy is still conjectural. Details of these cases are given below.

- $^{226}_{88}\text{RaC}$. Total disintegration energy taken as 3.17 MeV, rather than 3.78 MeV. Determination of E_β for $^{210}_{82}\text{Ac}$ as 1.17 ± 0.07 MeV (no γ -rays) by Hall and Templeton (1950) makes this the more probable assumption.
- $^{234}_{90}\text{MeTh}_2$. Total disintegration energy assumed 2.0 ± 0.2 MeV.
- $^{231}_{91}\text{Pa}$. Total disintegration energy assumed 0.58 MeV (Elliott, reported in *Nuclear Data*, NBS 499, 1950).
- $^{233}_{91}\text{Pa}$. UZ being the ground-state isomer, total disintegration energy taken as $2.32 - 0.39 = 1.93$ MeV (Bradt and Scherrer, 1945).
- $^{235}_{92}\text{U}$. Total disintegration energy assumed 0.60 ± 0.05 MeV (see Melander and Slätis, 1948).
- $^{237}_{93}\text{Np}$. Total disintegration energy assumed 0.68 MeV (see Fulbright, 1947).
- $^{239}_{94}\text{Bi}$. α -disintegration energy assumed 3.3 ± 0.3 MeV.

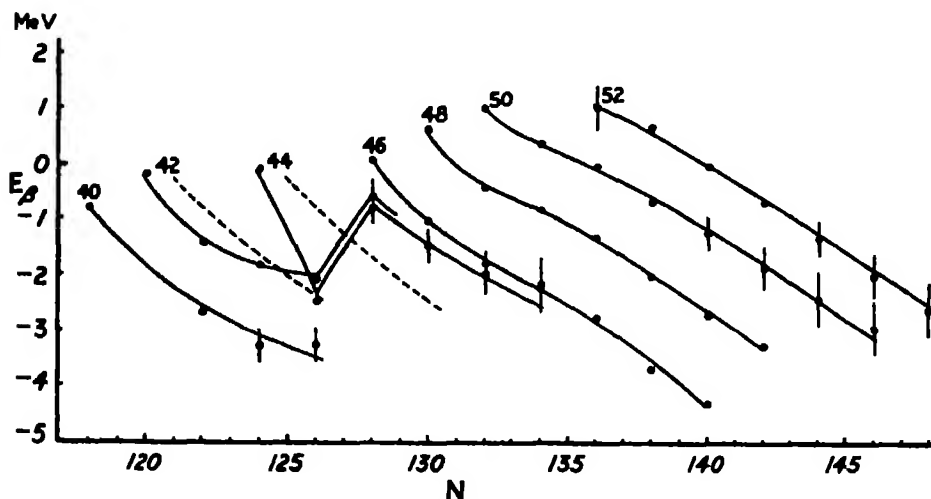


Fig. 4.— E_β as a function of N for species of even A and even Z . Isotopic numbers from 40 to 52 inclusive.

Reviewing figs. 2–4, it is obvious at first sight that discontinuities occur on each figure in the region of $N = 126$, as suggested by Suess, but closer inspection shows certain unexpected features. Thus on fig. 2 (odd-even and even-odd β -emitters) the discontinuity is evident only on the curve for isotopic number 43, the neighbouring curves for $A - 2Z = 41, 45$ showing no signs of irregularity over the same range of N .^{*} As is emphasized by comparison of the actual curve of E_β against N for $A - 2Z = 43$ with the dotted curves, the sense of the irregularity is that the β -disintegration energies of $^{205}_{81}\text{Tl}$, $^{207}_{82}\text{Pb}$ and $^{209}_{83}\text{Bi}$ are all low, and those of $^{213}_{85}\text{At}$ and $^{215}_{86}\text{Em}$ both high, in comparison with what might have been expected by simple interpolation. (The fact that all these species are β -stable—with the somewhat doubtful exception of

^{*} The curve for $A - 2Z = 41$ (Z even) shows a marked step between $^{172}_{78}\text{Os}$ ($N = 117$) and $^{174}_{78}\text{Pt}$ ($N = 119$), which would make it appear that $N = 118$ is a minor “magic” number.

$^{212}_{85}\text{At}$ —does not make this statement any less significant.) Abnormally high β -stability, therefore, is characteristic of the species ($Z = 82$, $N = 125$), ($Z = 83$, $N = 126$) and ($Z = 81$, $N = 124$)—here listed in order of decreasing abnormality, and abnormally low β -stability characteristic of ($Z = 85$, $N = 128$) and ($Z = 86$, $N = 129$). The β -stability of all other odd- A species represented in fig. 2 ($197 \leq A \leq 243$) is normal—except that it appears to be exceptionally low for $^{209}_{81}\text{Tl}$ ($Z = 81$, $N = 128$) and $^{211}_{81}\text{Tl}$ ($Z = 81$, $N = 130$). Possibly these last abnormalities are indicative of the occurrence on the curves for $A - 2Z = 47$ and 49 of discontinuities similar in form to that exhibited for $A - 2Z = 43$, but it is likely to be some considerable time before this possibility can be checked against experimental fact. Meanwhile the strict normality of the β -disintegration energies of all species having $A - 2Z = 45$ is the more remarkable. Obviously the incidence of shell-closure both at $Z = 82$ and at $N = 126$ complicates the position in respect of the species now under discussion, but the fact that the β -disintegration energy of $^{207}_{81}\text{Tl}$ ($Z = 81$, $N = 126$) is normal whereas that of $^{207}_{82}\text{Pb}$ ($Z = 82$, $N = 125$) is some 2.3 MeV low, that E_β for $^{209}_{82}\text{Pb}$ ($Z = 82$, $N = 127$) is normal and for $^{209}_{83}\text{Bi}$ ($Z = 83$, $N = 126$) some 1.5 MeV low, is at first sight surprising. It seems that it can only be explained on the assumption that the discontinuity in these cases is determined more by the change in Z than by the change in N .

On fig. 3 (odd-odd nuclear species) the only marked discontinuity appears on the curve for $A - 2Z = 42$. Here $^{212}_{85}\text{At}$ ($Z = 85$, $N = 127$) is abnormally β -unstable (to the extent of about 1 MeV) and $^{206}_{83}\text{Bi}$ ($Z = 83$, $N = 125$) abnormally β -stable (by about 0.6 MeV). A more complicated situation is revealed by fig. 4. Both for $A - 2Z = 42$ and for $A - 2Z = 44$ abrupt decreases in β -stability are indicated between $N = 126$ and $N = 128$ (as between $^{210}_{84}\text{Po}$ and $^{214}_{86}\text{Em}$ and between $^{208}_{82}\text{Pb}$ and $^{212}_{84}\text{Po}$, respectively)—that is over the range of neutron numbers which includes the “magic” number 126—but there is the added suggestion that for $A - 2Z = 42$ the species having $N = 120$, 122 and 124 are unusually β -stable, and for $A - 2Z = 44$ that those having $N = 130$, 132 and 134 are unusually β -unstable. This represents a wider range of influence of neutron shell-closure than is otherwise indicated. Beyond this remark, however, it does not appear profitable here to pursue these purely qualitative considerations farther.

THE α -DISINTEGRATION ENERGIES OF CERTAIN ISOTOPES OF LEAD AND BISMUTH

It has already been stated that, for the purpose of calculating certain energies of capture transformation, the energy of α -disintegration of $^{209}_{83}\text{Bi}$ has been assumed to be $3.3 \pm 0.3 \text{ MeV}$. This assumed value,

together with the known energies of β -disintegration of $^{209}_{82}\text{Pb}$ and $^{205}_{80}\text{Hg}$, leads to a value of 2.4 ± 0.4 MeV for the energy of α -disintegration of $^{209}_{82}\text{Pb}$. In a similar way, by the method of closed decay cycles, using the energies of β -disintegration plotted in figs. 2-4, we obtain energies of α -disintegration of $^{206}_{83}\text{Bi}$, $^{207}_{83}\text{Bi}$, $^{208}_{82}\text{Pb}$ and $^{207}_{82}\text{Pb}$. For these calculations the known energies of α -disintegration of $^{208}_{84}\text{Po}$ and $^{207}_{84}\text{Po}$ (Karraker, Ghiorso and Templeton, 1951) serve as starting material. Table III collects the values so obtained.

TABLE III

Species	$^{206}_{83}\text{Bi}$	$^{207}_{83}\text{Bi}$	$^{208}_{82}\text{Pb}$	$^{209}_{82}\text{Pb}$	$^{207}_{82}\text{Pb}$
α -disintegration energy (MeV)	(3.3 ± 0.3)	4.3 ± 0.1	4.25 ± 0.1	(2.4 ± 0.4)	2.0 ± 0.3

Species	$^{207}_{82}\text{Pb}$
α -disintegration energy (MeV)	1.17 ± 0.15

The surprising feature of the results given in this table is the smallness of the α -disintegration energies of the isotopes of lead (and particularly of $^{207}_{82}\text{Pb}$). This feature has not hitherto been detected or predicted by the systematizers (Pryce, 1950).

REFERENCES TO LITERATURE

- ALBOUY and TELLAC, 1951. *O.R. Acad. Sci.*, Paris, CCXXII, 326.
 ALBUROER and FRIEDLANDER, 1951. *Phys. Rev.*, LXXXI, 523.
 BARTON, GHIORSO and PERLMAN, 1951. *Ibid.*, LXXXII, 13.
 BRADT and SCHERRER, 1945. *Helv. Phys. Acta*, XVIII, 405.
 ———, 1946. *Ibid.*, XIX, 307.
 CRANE, GHIORSO and PERLMAN, 1949. Unpublished.
 EVANS, 1950. *Proc. Phys. Soc.*, A, LXII, 575.
 FEATHER and RICHARDSON, 1948. *Ibid.*, LXI, 452.
 FULBRIGHT, 1947. *MDDC* 753.
 HALL and TEMPLETON, 1950. *UCRL* 057.
 HOWLAND, TEMPLETON and PERLMAN, 1947. *Phys. Rev.*, LXXI, 552.
 JAFFEY and MAGNUSON, 1949. *AECD* 2636.
 JAMES, FLORIN, HOPKINS and GHIORSO, 1948. *NNES* 14B, 22.8.
 ———, THOMPSON and HOPKINS, 1948. *NNES* 14B, 22.16.
 KARRAKER, GHIORSO and TEMPLETON, 1951. *UCRL* 1091.
 ——— and TEMPLETON, 1950. *UCRL* 640.
 KNIGHT and MACKLIN, 1949. *Phys. Rev.*, LXXV, 34.
 KONOFINSKI, 1943. *Rev. Mod. Phys.*, xv, 209.
 LUTZ, POOL and KURBATOV, 1944. *Phys. Rev.*, LXV, 61.
 MAGNUSON, THOMPSON and SEABORG, 1950. *Ibid.*, LXXVIII, 363.
 MAURER and RAMM, 1942. *Zeits. Physik*, CXIX, 602.
 MELANDER and SLÄTIS, 1948. *Phys. Rev.*, LXXIV, 709.
 MONSKOWNKI, 1951. *Ibid.*, LXXXII, 35.
 NEUMANN and PERLMAN, 1950. *Ibid.*, LXXVIII, 191.
 ———, 1951. *Ibid.*, LXXXI, 958.

256 *N. Feather, The Sargent Diagram for the Electron-capture Process*

- OSBORNE, THOMPSON and VAN WINKLE, 1946. *NNES* 14B, 19.11.
PRYCE, 1950. *Proc. Phys. Soc., A*, LXIII, 692.
RENNARD, 1949. *C.R. Acad. Sci., Paris*, CCXXVIII, 387.
SAHA and SAHA, 1946. *Trans. Nat. Inst. Sci., India*, 11, 193.
SARGENT, 1933. *Proc. Roy. Soc., A*, CXXXIX, 659.
—, 1939. *Can. Journ. Research.*, XVII, 82.
SEABORG, JAMES and MORGAN, 1948. *NNES* 14B, 22.1.
SLATIS and SIEGBAHN, 1949. *Arkiv Mat. Astron. Fysik*, XXXVIA, 21.
STUDIER, 1947. *AEC* 2444.
SULLIVAN, KOHMAN and SWARTOUT, 1945. Plutonium Project Report, *HEW* 3, 1635.
SUNYAE, ALBURGER, FRIEDLANDER, GOLDBABER and SCHARFF-GOLDBABER, 1950. *Phys. Rev.*, LXXVIII, 326.
TEMPLETON, 1950. *UORL* 616.
—, HOWLAND and PERLMAN, 1947. *Phys. Rev.*, LXXII, 758.
THOMPSON, 1949. *Ibid.*, LXXVI, 319.

(Issued separately February 9, 1952)

XVIII.—The Elementary Proof of the Prime Number Theorem.*

By E. M. Wright, University of Aberdeen.

(MS. received June 4, 1951. Read November 5, 1951)

SYNOPSIS

The author presents a modification of the recently discovered "elementary" proof of the Prime Number Theorem. Nothing is assumed from the theory of numbers except the Fundamental Theorem of Arithmetic. In the second part of the proof the elements of the integral calculus are used to make clearer the basic ideas on which this part depends.

1. The Prime Number Theorem asserts that $\pi(x)$, the number of primes less than or equal to x , is asymptotic to $x/\log x$ as $x \rightarrow \infty$. This was first proved (independently) by Hadamard (1896) and de la Vallée Poussin (1896). Many authors shortened and simplified the proof, but it was not until very recently that Erdős and Selberg (Erdős, 1949) and Selberg (1949) gave proofs which are "elementary" (in the technical logical sense). Each of these proofs depends on a very striking inequality due to Selberg and on a few well-known theorems in the elementary theory of primes. Van der Corput (1949) presented the Erdős-Selberg form of the proof *ab initio*, giving classical proofs of the known results which he requires. Here I present a modification of Selberg's proof, in which I assume nothing from the theory of numbers, except the Fundamental Theorem of Arithmetic that every integer $n > 1$ is the product of primes, i.e.,

$$n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k} \quad (p_1 < p_2 < \dots < p_k), \quad (1.1)$$

in just one way. (This is, of course, implicitly assumed by all the authors mentioned.) But, apart from certain trivialities, I do not use the classical proofs of the known results required (especially (4.3) and (4.4) below), but obtain these as by-products of the main argument. From "analysis" I assume the well-known result

$$\sum_{n \leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right) \quad (x \geq 1) \quad (1.2)$$

but less would suffice (see § 10).

PROOF OF SELBERG'S INEQUALITY

2. In what follows m, n, d, k, \dots denote positive integers, p (with or without a suffix) always a prime number and x and y any real numbers not less than 1. The $O(\)$ notation refers to the passage of x to infinity.

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

As usual, $[x]$ is the greatest integer not exceeding x , and $d|n$ denotes that $n = dk$ for some integral k . The Möbius function $\mu(n)$ is defined by:

$$\mu(1) = 1, \quad \mu(p_1 p_2 \dots p_k) = (-1)^k,$$

provided all the p are different, and $\mu(n) = 0$ if n is divisible by any p^2 . We write also

$$\Lambda(n) = \log p \quad (n = p^a), \quad \Lambda(n) = 0$$

for all other n and

$$\psi(x) = \sum_{n \leq x} \Lambda(n).$$

If $n > 1$ has the form (1.1), the only divisors d of n for which $\mu(d) \neq 0$ are $1, p_1, \dots, p_k, p_i p_j (i \neq j), p_i p_j p_l, \dots$, etc., and so

$$\begin{aligned} \sum_{d|n} \mu(d) &= \mu(1) + \sum \mu(p_i) + \sum_{i \neq j} \mu(p_i p_j) + \dots \\ &= 1 - \binom{k}{1} + \binom{k}{2} - \binom{k}{3} + \dots = (1 - 1)^k = 0. \end{aligned} \quad (2.1)$$

On the other hand,

$$\sum_{d|1} \mu(d) = \mu(1) = 1. \quad (2.2)$$

Also,

$$\sum_{d|n} \Lambda(d) = \sum_{i=1}^k \sum_{a=1}^{a_i} \Lambda(p_i^a) = \sum_{i=1}^k a_i \log p_i = \log n. \quad (2.3)$$

By (2.1), (2.2) and (2.3),

$$\begin{aligned} - \sum_{d|n} \mu(d) \log d &= \sum_{d|n} \mu(d) \log \left(\frac{n}{d} \right) = \sum_{d|n} \mu(d) \sum_{h|\frac{n}{d}} \Lambda(h) \\ &= \sum_{h|n} \Lambda(h) \sum_{d|\frac{n}{h}} \mu(d) = \Lambda(n). \end{aligned} \quad (2.4)$$

Hence

$$\begin{aligned} \sum_{d|n} \Lambda(d) \Lambda \left(\frac{n}{d} \right) &= - \sum_{d|n} \Lambda(d) \sum_{h|\frac{n}{d}} \mu(h) \log h \\ &= - \sum_{h|n} \mu(h) \log h \sum_{d|\frac{n}{h}} \Lambda(d) = - \sum_{h|n} \mu(h) \log h \log \left(\frac{n}{h} \right) \\ &= \Lambda(n) \log n + \sum_{d|n} \mu(d) \log^2 d. \end{aligned} \quad (2.5)$$

Hence, if $n > 1$,

$$\begin{aligned} \sum_{d|n} \mu(d) \log^2 \left(\frac{x}{d} \right) &= -2 \log x \sum_{d|n} \mu(d) \log d + \sum_{d|n} \mu(d) \log^2 d \\ &= 2\Lambda(n) \log x - \Lambda(n) \log n + \sum_{d|n} \Lambda(d) \Lambda \left(\frac{n}{d} \right), \end{aligned}$$

while

$$\sum_{d|1} \mu(d) \log^2 \left(\frac{x}{d} \right) = \log^2 x.$$

Hence, if we write

$$S(x) = \sum_{n \leq x} \sum_{d|n} \mu(d) \log^2 \left(\frac{x}{d} \right),$$

we have

$$\begin{aligned} S(x) &= \log^2 x + 2\psi(x) \log x - \sum_{n \leq x} \Lambda(n) \log n + \sum_{d \leq x} \Lambda(d) \Lambda(x/d) \\ &= \log^2 x + \psi(x) \log x + \sum_{n \leq x} \Lambda(n) \log \left(\frac{x}{n} \right) + \sum_{m \leq x} \Lambda(m) \Lambda(n). \end{aligned} \quad (2.6)$$

Since no term in the last expression can be negative, we have

$$\psi(x) \log x \leq S(x). \quad (2.7)$$

3. If $xe^{-k} \leq n < xe^{1-k}$, we have

$$k-1 < \log \left(\frac{x}{n} \right) \leq k$$

and so, if $h > 0$,

$$\sum_{n \leq x} \log^h \left(\frac{x}{n} \right) \leq \sum_{k \geq 1} k^h \sum_{xe^{-k} \leq n < xe^{1-k}} 1 < x \sum_{k=1}^{\infty} k^h e^{1-k} = O(x). \quad (3.1)$$

If we put $h = 1$, this gives us

$$\sum_{n \leq x} \log n = [x] \log x + O(x) = x \log x + O(x). \quad (3.2)$$

Among the numbers $1, 2, 3, \dots, [x]$, there are just $[x/p]$ multiples of p , just $[x/p^2]$ multiples of p^2 , and so on. Hence $[x]!$ is divisible by p just j times, where

$$j = \left[\frac{x}{p} \right] + \left[\frac{x}{p^2} \right] + \left[\frac{x}{p^3} \right] + \dots$$

Hence

$$\sum_{n \leq x} \log n = \log([x]!) = \sum_p \log p \left(\left[\frac{x}{p} \right] + \left[\frac{x}{p^2} \right] + \dots \right) = \sum_{n \leq x} \left[\frac{x}{n} \right] \Lambda(n).$$

If we remove the square brackets we introduce an error of at most

$$\sum_{n \leq x} \Lambda(n) = \psi(x).$$

Hence, by (3.2),

$$x \sum_{n \leq x} \frac{\Lambda(n)}{n} = x \log x + O(x) + O(\psi(x)). \quad (3.3)$$

4. By (2.1) and (2.2),

$$\begin{aligned} S(x) - \gamma^2 &= \sum_{n \leq x} \sum_{d|n} \mu(d) \left\{ \log^2 \left(\frac{x}{d} \right) - \gamma^2 \right\} \\ &= \sum_{d \leq x} \mu(d) \left[\frac{x}{d} \right] \left\{ \log^2 \left(\frac{x}{d} \right) - \gamma^2 \right\}, \end{aligned}$$

since there are $[x/d]$ multiples of d which are $\leq x$. If we remove the square brackets the error is at most

$$\sum_{d \leq x} \left\{ \log^2 \left(\frac{x}{d} \right) + \gamma^2 \right\} = O(x)$$

by (3.1) with $h = 2$. Hence, by (1.2),

$$\begin{aligned} S(x) &= x \sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \log^2 \left(\frac{x}{d} \right) - \gamma^2 \right\} + O(x) \\ &= x \sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \log \left(\frac{x}{d} \right) - \gamma \right\} \left\{ \sum_{k \leq \frac{x}{d}} \frac{1}{k} + O \left(\frac{d}{x} \right) \right\} + O(x) \\ &= x \sum_{d \leq x} \frac{\mu(d)}{dk} \left\{ \log \left(\frac{x}{d} \right) - \gamma \right\} + O(x) \\ &= x \sum_{n \leq x} \frac{1}{n} \sum_{d|n} \mu(d) \{ \log(x/d) - \gamma \} + O(x) \\ &= x \log x + x \sum_{n \leq x} \frac{\Lambda(n)}{n} + O(x) \end{aligned} \quad (4.1)$$

by (2.1), (2.2) and (2.4). Hence, by (3.3),

$$S(x) = 2x \log x + O(x) + O\{\psi(x)\}. \quad (4.2)$$

By (2.7) and (4.2), we have

$$\psi(x) = O(x) \quad (4.3)$$

and so (3.3) gives us

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \log x + O(1). \quad (4.4)$$

Also, as in (3.1),

$$\begin{aligned} \sum_{n \leq x} \Lambda(n) \log \left(\frac{x}{n} \right) &\leq \sum_{k \geq 1} k \sum_{x/k \leq n < x^{1-k}} \Lambda(n) \leq \sum_{k \geq 1} k \psi(xe^{1-k}) \\ &= O \left(x \sum_{k \geq 1} ke^{1-k} \right) = O(x). \end{aligned} \quad (4.5)$$

Combining (2.6) and (4.2) and using (4.3) and (4.5), we obtain

$$\psi(x) \log x + \sum_{mn \leq x} \Lambda(m) \Lambda(n) = 2x \log x + O(x). \quad (4.6)$$

From this we have the following modified form of Selberg's inequality :

$$\psi(x) \log x + \sum_{n \leq x} \Lambda(n) \psi \left(\frac{x}{n} \right) = 2x \log x + O(x). \quad (4.7)$$

Hence, by (4.5),

$$\sum_{n \leq x} \Lambda(n) \log n + \sum_{mn \leq x} \Lambda(m) \Lambda(n) = 2x \log x + O(x). \quad (4.8)$$

PROOF THAT $\psi(x) \sim x$

5. We now write $R(x) = \psi(x) - x$ in (4.7). Using (4.4) we have

$$R(x) \log x + \sum_{n \leq x} \Lambda(n) R\left(\frac{x}{n}\right) = O(x).$$

If we replace n by m and x by x/n , this becomes

$$R\left(\frac{x}{n}\right) \log\left(\frac{x}{n}\right) + \sum_{m \leq x/n} \Lambda(m) R\left(\frac{x}{mn}\right) = O\left(\frac{x}{n}\right).$$

Hence

$$\begin{aligned} \log x \left\{ R(x) \log x + \sum_{n \leq x} \Lambda(n) R\left(\frac{x}{n}\right) \right\} \\ - \sum_{n \leq x} \Lambda(n) \left\{ R\left(\frac{x}{n}\right) \log\left(\frac{x}{n}\right) + \sum_{m \leq x/n} \Lambda(m) R\left(\frac{x}{mn}\right) \right\} \\ = O(x \log x) + O\left(x \sum_{n \leq x} \frac{\Lambda(n)}{n}\right) = O(x \log x) \end{aligned}$$

and this reduces to

$$R(x) \log^2 x = - \sum_{n \leq x} R\left(\frac{x}{n}\right) \Lambda(n) \log n + \sum_{mn \leq x} R\left(\frac{x}{mn}\right) \Lambda(m) \Lambda(n) + O(x \log x),$$

whence

$$|R(x)| \log^2 x \leq \sum_{n \leq x} a_n \left| R\left(\frac{x}{n}\right) \right| + O(x \log x), \quad (5.1)$$

where

$$a_n = \Lambda(n) \log n + \sum_{dk=n} \Lambda(d) \Lambda(k)$$

and, by (4.8),

$$A(x) = \sum_{n \leq x} a_n = 2x \log x + O(x). \quad (5.2)$$

We now approximate to the sum on the right-hand side of (5.1) by means of an integral. If $y > y' > 0$, we have

$$\begin{aligned} ||R(y)| - |R(y')|| \leq |R(y) - R(y')| \leq \psi(y) - \psi(y') + (y - y') \\ = f(y) - f(y'), \end{aligned} \quad (5.3)$$

where

$$0 < f(y) = \psi(y) + y = O(y)$$

by (4.3) and $f(y)$ increases steadily with y . Also

$$\sum_{n \leq x-1} n \left\{ f\left(\frac{x}{n}\right) - f\left(\frac{x}{n+1}\right) \right\} = \sum_{n \leq x} f\left(\frac{x}{n}\right) - [x] f\left(\frac{x}{[x]}\right) = O(x \log x). \quad (5.4)$$

Now let us write

$$b_1 = a_1 = 0, \quad b_n = a_n - 2 \int_{n-1}^n \log t \, dt \quad (n > 1)$$

and

$$B(x) = \sum_{n \leq x} b_n = A(x) - 2 \int_1^{[x]} \log t \, dt = O(x) \quad (5.5)$$

by (5.2). Hence

$$\begin{aligned} \sum_{n \leq x} b_n \left| R\left(\frac{x}{n}\right) \right| &= \sum_{n \leq x} \{B(n) - B(n-1)\} \left| R\left(\frac{x}{n}\right) \right| \\ &= \sum_{n \leq x-1} B(n) \left\{ \left| R\left(\frac{x}{n}\right) \right| - \left| R\left(\frac{x}{n+1}\right) \right| \right\} + B(x) \left| R\left(\frac{x}{[x]}\right) \right| \\ &= O\left(\sum_{n \leq x-1} n \left\{ f\left(\frac{x}{n}\right) - f\left(\frac{x}{n+1}\right) \right\} \right) + O(x) \\ &= O(x \log x) \end{aligned}$$

by (5.3), (5.4) and (5.5). Hence

$$\sum_{n \leq x} a_n \left| R\left(\frac{x}{n}\right) \right| = 2 \sum_{z \leq n \leq x} \left| R\left(\frac{x}{n}\right) \right| \int_{n-1}^n \log t \, dt + O(x \log x). \quad (5.6)$$

Again, for $n \geq 2$,

$$\begin{aligned} \left| \left| R\left(\frac{x}{n}\right) \right| \int_{n-1}^n \log t \, dt - \int_{n-1}^n \left| R\left(\frac{x}{t}\right) \right| \log t \, dt \right| \\ \leq \int_{n-1}^n \log t \left\{ f\left(\frac{x}{t}\right) - f\left(\frac{x}{n}\right) \right\} dt \leq (n-1) \left\{ f\left(\frac{x}{n-1}\right) - f\left(\frac{x}{n}\right) \right\} \end{aligned}$$

and so

$$\begin{aligned} \sum_{z \leq n \leq x} \left| R\left(\frac{x}{n}\right) \right| \int_{n-1}^n \log t \, dt - \int_1^x \left| R\left(\frac{x}{t}\right) \right| \log t \, dt \\ = O\left(\sum_{n \leq x} (n-1) \left\{ f\left(\frac{x}{n-1}\right) - f\left(\frac{x}{n}\right) \right\} \right) + O\left(\int_{[x]}^x \left| R\left(\frac{x}{t}\right) \right| \log t \, dt \right) \\ = O(x \log x) \end{aligned} \quad (5.7)$$

by (5.4).

By (5.1), (5.6) and (5.7), we have

$$|R(x)| \log^2 x \leq 2 \int_1^x \left| R\left(\frac{x}{t}\right) \right| \log t \, dt + O(x \log x). \quad (5.8)$$

We write

$$V(\xi) = e^{-\xi} R(e\xi) = e^{-\xi} \psi(e\xi) - 1 = e^{-\xi} \left(\sum_{n \leq e\xi} \Lambda(n) \right) - 1. \quad (5.9)$$

If we put $x = e\xi$ and $t = x e^{-\eta}$, we have

$$\begin{aligned} \int_1^x \left| R\left(\frac{x}{t}\right) \right| \log t \, dt &= x \int_1^{\xi} |V(\eta)| (\xi - \eta) d\eta = x \int_0^{\xi} |V(\eta)| \left(\int_{\eta}^{\xi} d\zeta \right) d\eta \\ &= x \int_0^{\xi} d\zeta \int_0^{\zeta} |V(\eta)| d\eta \end{aligned}$$

and (5.8) becomes

$$|V(\xi)| \xi^{\alpha} \leq 2 \int_0^{\xi} d\zeta \int_0^{\zeta} |V(\eta)| d\eta + O(\xi). \quad (5.10)$$

6. We have

$$\begin{aligned} \int_1^x \frac{\psi(t)}{t^2} dt &= \int_1^x \left\{ \sum_{n \leq t} \Lambda(n) \right\} \frac{dt}{t^2} = \sum_{n \leq x} \Lambda(n) \int_n^x \frac{dt}{t^2} \\ &= \sum_{n \leq x} \Lambda(n) \left(\frac{1}{n} - \frac{1}{x} \right) = \log x + O(1). \end{aligned}$$

by (4.3) and (4.4). Putting $t = e^{\eta}$, $x = e^{\xi}$, we find that

$$\int_0^{\xi} V(\eta) d\eta = \int_1^x \left(\frac{\psi(t)}{t^2} - \frac{1}{t} \right) dt = O(1).$$

Hence there is a fixed number A such that, for all $\xi_2 > \xi_1 > 0$,

$$\left| \int_{\xi_1}^{\xi_2} V(\eta) d\eta \right| < A. \quad (6.1)$$

Next let $y > y_0$. Writing (4.6) with $x = y$ and again with $x = y_0$ and subtracting, we obtain

$$\begin{aligned} 0 &\leq \psi(y) \log y - \psi(y_0) \log y_0 \\ &= 2y \log y - 2y_0 \log y_0 - \sum_{y_0 < m \leq y} \Lambda(m) \Lambda(n) + O(y) \\ &\leq 2(y \log y - y_0 \log y_0) + O(y), \end{aligned}$$

since $\Lambda(n) \geq 0$ for all n . Hence

$$|R(y) \log y - R(y_0) \log y_0| \leq y \log y - y_0 \log y_0 + O(y). \quad (6.2)$$

Now let us write $y = e^{\eta}$, $y_0 = e^{\eta_0}$, $\tau = \eta - \eta_0 > 0$ and suppose that $V(\eta_0) = 0$. We have from (6.2)

$$|V(\eta)| \leq 1 - \frac{\eta_0}{\eta} e^{\eta_0 - \eta} + O\left(\frac{1}{\eta}\right) = 1 - e^{-\tau} + \frac{\tau e^{-\tau}}{\eta} + O\left(\frac{1}{\eta}\right) < \tau + O\left(\frac{1}{\eta}\right),$$

since

$$\tau e^{-\tau} < 1, \quad 1 - e^{-\tau} < \tau.$$

Hence

$$|V(\eta)| < \eta - \eta_0 + O(\eta^{-1}), \quad (6.3)$$

provided $\eta > \eta_0$ and $V(\eta_0) = 0$.

7. Our object is to prove that $V(\xi) \rightarrow 0$ as $\xi \rightarrow \infty$. By (4.3) and (5.9), $V(\xi)$ is bounded and so there is a number α such that*

$$|V(\xi)| \leq \alpha + o(1). \quad (7.1)$$

* We use $o(1)$ to denote a function of ξ (or η or ζ as the case may be) which tends to zero as $\xi \rightarrow \infty$ and which is bounded for all $\xi > 0$. Also $o(\xi) = \xi o(1)$, $o(\xi^2) = \xi^2 o(1)$. Hence

$$\int_0^{\xi} o(1) d\eta = o(\xi), \quad \int_0^{\xi} o(\eta) d\eta = o(\xi^2).$$

If $\alpha = 0$, we have nothing to prove. Hence let us suppose that $\alpha > 0$. We take $\zeta > 0$ and write

$$\beta = \frac{3\alpha^2 + 4A}{2\alpha} > \alpha.$$

From the definition (5.9) of $V(\xi)$, it is clear that $V(\xi)$ decreases continuously as ξ increases except at points of discontinuity (*viz.*, those at which $\xi = \log p^m$) at which $V(\xi)$ increases. Hence $V(\xi)$ can only change from positive to negative by passing through a zero. Hence, for a given $\zeta > 0$, either there is an η_0 such that

$$\zeta \leq \eta_0 \leq \zeta + \beta - \alpha, \quad V(\eta_0) = 0$$

or $V(\eta)$ cannot change sign more than once in the interval

$$\zeta \leq \eta \leq \zeta + \beta - \alpha. \quad (7.2)$$

In the former case we have, by (6.3) and (7.1),

$$\begin{aligned} \int_{\zeta}^{\zeta+\beta} |V(\eta)| d\eta &= \int_{\zeta}^{\eta_0} + \int_{\eta_0}^{\eta_0+\alpha} + \int_{\eta_0+\alpha}^{\zeta+\beta} |V(\eta)| d\eta \\ &\leq (\eta_0 - \zeta)\alpha + \int_0^{\alpha} r dr + (\zeta + \beta - \eta_0 - \alpha)\alpha + o(1) \\ &= \alpha\beta - \frac{1}{2}\alpha^2 + o(1) = \alpha'\beta + o(1), \end{aligned}$$

where

$$\alpha' = \alpha \left(1 - \frac{\alpha}{2\beta}\right) = \alpha \left(1 - \frac{\alpha^2}{3\alpha^2 + 4A}\right). \quad (7.3)$$

In the latter case, either $V(\eta)$ does not change sign in the interval (7.2), so that

$$\int_{\zeta}^{\zeta+\beta-\alpha} |V(\eta)| d\eta = \left| \int_{\zeta}^{\zeta+\beta-\alpha} V(\eta) d\eta \right| < A,$$

or $V(\eta)$ changes sign just once in the interval, say at $\eta = \eta_1$, when

$$\int_{\zeta}^{\zeta+\beta-\alpha} |V(\eta)| d\eta = \left| \int_{\zeta}^{\eta_1} V(\eta) d\eta \right| + \left| \int_{\eta_1}^{\zeta+\beta-\alpha} V(\eta) d\eta \right| < 2A.$$

In either case, by (7.1),

$$\int_{\zeta}^{\zeta+\beta} |V(\eta)| d\eta < 2A + \int_{\zeta+\beta-\alpha}^{\zeta+\beta} |V(\eta)| d\eta < 2A + \alpha^2 + o(1) = \alpha''\beta + o(1),$$

where
$$\alpha'' = \frac{2A + \alpha^2}{\beta} = \alpha \left(\frac{4A + 2\alpha^2}{4A + 3\alpha^2} \right) = \alpha \left(1 - \frac{\alpha}{2\beta} \right) = \alpha'.$$

Hence, if $M = [\xi/\beta]$,

$$\begin{aligned} \int_0^{\xi} |V(\eta)| d\eta &= \sum_{m=0}^{M-1} \int_{m\beta}^{m\beta+\beta} |V(\eta)| d\eta + \int_{M\beta}^{\xi} |V(\eta)| d\eta \\ &\leq \alpha'\beta M + o(M) + O(1) = \alpha'\xi + o(\xi). \end{aligned}$$

Hence, by (5.10),

$$|V(\xi)|\xi^2 \leq 2 \int_0^\xi \{\alpha'\zeta + o(\zeta)\} d\zeta + O(\xi) = \alpha'\xi^2 + o(\xi^2)$$

and so

$$|V(\xi)| \leq \alpha' + o(1). \quad (7.4)$$

8. We can complete the proof in two alternative ways. We may take

$$\alpha = \overline{\lim_{\xi \rightarrow \infty}} |V(\xi)|.$$

If $\alpha > 0$, it follows from (7.4) that $\alpha \leq \alpha'$ and from (7.3) that $\alpha' < \alpha$. This is a contradiction and so $\alpha = 0$, that is $V(\xi) \rightarrow 0$ and

$$\psi(x) \sim x. \quad (8.1)$$

If we wish, we can avoid an indirect proof and the idea of an upper limit as follows. We know that (7.1) is true when α is equal to some $\bar{\alpha}$ and we can take $A \geq 3\bar{\alpha}^2$. We write $\alpha_n = \sqrt{(5A/n)}$, so that $\alpha_{15} \geq \bar{\alpha}$ and (7.1) is true for $\alpha = \alpha_{15}$. But, if $n \geq 15$ and $\alpha = \alpha_n \leq \alpha_{15} = \sqrt{(A/3)}$, we have

$$\alpha' = \alpha_n \left(1 - \frac{\alpha_n^2}{3\alpha_n^2 + 4A}\right) \leq \alpha_n \left(1 - \frac{\alpha_n^2}{5A}\right) = \sqrt{\frac{5A}{n}} \left(\frac{n-1}{n}\right) < \alpha_{n+1}.$$

Hence (7.1) is true for $\alpha = \alpha_{15}, \alpha_{17}, \alpha_{18}, \dots$. Since $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$, it follows that $V(\xi) \rightarrow 0$ as $\xi \rightarrow \infty$.

PROOF OF THE PRIME NUMBER THEOREM

9. We write as usual

$$\pi(x) = \sum_{p \leq x} 1 = O(x).$$

We have

$$\begin{aligned} \psi(x) &= \sum_{n \leq x} \Lambda(n) = \sum_{p^a \leq x} \log p \leq \log x \sum_{p^a \leq x} 1 \\ &= \log x \sum_{a \leq \left[\frac{\log x}{\log 2}\right]} \pi(x^{1/a}) = \pi(x) \log x + O(x^{\frac{1}{2}} \log^2 x) \end{aligned}$$

and so

$$\pi(x) \log x \geq x + o(x).$$

Next, if $X = x/\log x$,

$$\psi(x) \geq \sum_{x < p \leq x} \log p \geq \{\pi(x) - \pi(X)\} \log X,$$

$$\pi(x) \leq \pi(X) + \frac{\psi(x)}{\log X} = O(X) + \frac{O(x)}{\log X} = O\left(\frac{x}{\log x}\right),$$

$$\pi(X) = O\left(\frac{X}{\log X}\right) = O\left(\frac{x}{\log^2 x}\right),$$

$$\pi(x) \log x \leq \pi(X) \log x + \frac{\psi(x) \log x}{\log X} \sim x.$$

Hence

$$\pi(x) \sim \frac{x}{\log x}$$

as $x \rightarrow \infty$, which is the Prime Number Theorem.

REMARKS

10. It is enough for our purpose if the error term in (1.2) is replaced by $O(x^{-\delta})$ for some fixed $\delta > 0$. In (4.1) we have then an error of order

$$\begin{aligned} \sum_{d \leq x} \left(\frac{x}{d}\right)^{1-\delta} \left\{ \log \left(\frac{x}{d}\right) + \gamma \right\} \\ \leq \sum_{k=1}^{\infty} \sum_{x^{-\delta} \leq d < x^{1-\delta}} \left(\frac{x}{d}\right)^{1-\delta} \left\{ \log \left(\frac{x}{d}\right) + \gamma \right\} \\ \leq \sum_{k=1}^{\infty} (k + \gamma) e^{k(1-\delta)} x e^{1-k} = ex \sum_{k=1}^{\infty} (k + \gamma) e^{-k\delta} = O(x). \end{aligned}$$

If the error term in (1.2) is replaced by $o(1/\log x)$, however, we need to modify our argument somewhat tediously. In general, $O(x)$ is replaced by $o(x \log x)$ and, for example, $O(\xi)$ in (5.10) becomes $o(\xi^2)$. This suffices to obtain the prime number theorem, of course.

11. Selberg's inequality in its original form is

$$\vartheta(x) \log x + \sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) \log p = 2x \log x + O(x), \quad (11.1)$$

where

$$\vartheta(x) = \sum_{p \leq x} \log p.$$

This can be deduced simply from (4.7).

12. The use of the elements of the integral calculus in §§ 5-7 makes my presentation of the proof less "elementary" (in the technical logical sense) than Selberg's (1949), which uses the theory of series only. Selberg's proof could be applied to my (4.7) in just the same way as to (11.1). But the use of the calculus makes unnecessary certain complexities of detail and shows more clearly the fundamental ideas (which are, of course, essentially Selberg's).

REFERENCES TO LITERATURE

- ERDŐS, P., 1949. "On a New Method in Elementary Number Theory which Leads to an Elementary Proof of the Prime Number Theorem", *Proc. Nat. Acad. Sci.*, xxxv, 374-384.

- HADAMARD, J., 1896. "Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques", *Bull. Soc. Math. France*, xxiv, 199-220.
- SELEGNY, A., 1949. "An Elementary Proof of the Prime-number Theorem", *Annals of Math.*, L, 305-313.
- DE LA VALLÉE POUSSIN, CH., 1896. "Recherches analytiques sur la théorie des nombres premiers", *Ann. Soc. Sc. Bruxelles*, xxi B, 183-256.
- VAN DER CORPUT, J. G., 1949. "Démonstration élémentaire du théorème sur la distribution des nombres premiers", *Scriptum* 1 (Mathematisch Centrum, Amsterdam), 1-32.

(Issued separately May 30, 1952)

XIX.—A Generalization of the Classical Random-walk problem, and a Simple Model of Brownian Motion Based Thereon.

By G. Klein, Ph.D., Birkbeck College, London, now at Université Libre de Bruxelles*. Communicated by Dr. R. FÜRTH.

(MS. received September 18, 1951. Read December 3, 1951)

SYNOPSIS

Suggested by the analogy between the classical one-dimensional random-walk and the approximate (diffusion) theory of Brownian motion, a generalization of the random-walk is proposed to serve as a model for the more accurate description of the phenomenon. Using the methods of the calculus of finite differences, some general results are obtained concerning averages based on a time-varying bivariate discrete probability distribution in which the variates stand in the particular relation of "position" and "velocity." These are applied to the special cases of Brownian motion from initial thermal equilibrium, and from arbitrary initial kinetic energy. In the latter case the model describes accurately quantized Brownian motion of two energy states, one of zero energy.

INTRODUCTION

The classical random-walk problem, proposed by Pearson, has been thoroughly treated by Rayleigh. Independently, Smoluchowski used equivalent mathematical methods to describe Brownian motion, and this phenomenon has been linked with the random-walk problem ever since. The latest contribution has come from Kac, who generalized the problem so as to make it correspond to Brownian motion of an elastically bound particle, and to include boundary problems. The connection and treatment given so far hold however only for the "approximate" theory of Brownian motion, that is, for times $t \gg 1/\gamma$, where γ is the resistance coefficient per unit mass of the medium. Under these conditions for a "free" particle, Einstein's formula holds,

$$\overline{x^2} = 2 \frac{kT}{m\gamma} t, \quad (1)$$

which is derived from statistical mechanics. Here k is Boltzmann's constant, T the absolute temperature of the medium, m the mass of the Brownian particle, and $\overline{x^2}$ its average squared displacement during an interval, t .

A complete and more accurate description of Brownian motion can follow only from a detailed kinetic theory, but if mean squared displacement formulae only are required, these can be deduced by various simpler

* Most of the contents of this paper forms part of the author's Ph.D. Thesis, University of London, 1951.

methods as only the principle of equipartition of energy is involved (e.g., by an accurate integration of Langevin's pseudo-equation of motion of a Brownian particle as given by Ornstein), and it may be taken as established that the formulae replacing (1) are

$$\overline{x^2} = 2 \frac{kT}{m\gamma} t - 2 \frac{kT}{m\gamma^2} (1 - e^{-\gamma t}), \quad (2)$$

for conservative diffusion, that is, for a particle initially already in thermal equilibrium with the medium, and

$$\overline{x^2} = 2 \frac{kT}{m\gamma} t + \frac{kT}{m\gamma^2} (1 - e^{-2\gamma t}) - 4 \frac{kT}{m\gamma^2} (1 - e^{-\gamma t}) + \frac{u_0^2}{\gamma^2} (1 - e^{-\gamma t})^2, \quad (3)$$

for one starting with arbitrary kinetic energy $\frac{1}{2} m u_0^2$ (Chandrasekhar, 1943).

In order not to sever the connection with the random-walk problem for the "accurate" theory, the simplicity of equal discrete random displacements had to give way to the assumption of an appropriate continuous probability distribution for the magnitude of these displacements, and the problem, like the classical random-walk, becomes a special case of the theory of Markoff processes (Chandrasekhar, 1943).

Now it will be shown that the "accurate" results, (2) and (3), can also be obtained from a model more akin to the classical random-walk if, following Fürth, one introduces a certain "persistence," suggested by the physical situation. Mathematically this is equivalent to a discrete velocity distribution of only three variates, instead of the continuous one just mentioned. Although in some respects such a model must appear artificial it is of intrinsic interest, and it seems to be the simplest that is mathematically precise and at the same time shows the essential features of the actual physical phenomenon.

GENERAL THEORY

Let $f(n, m; N)$ be a discrete bivariate distribution function in n and m which varies in time, N signifying the *time* instant considered. If p 's are probabilities of transition during the N^{th} interval, the change in the distribution can be described completely by equations

$$f(n, m; N+1) = \sum_{n'} \sum_{m'} p_{n' \rightarrow n, m' \rightarrow m; N} f(n', m'; N). \quad (4)$$

Now let the variates n and m be related in a special manner, such that

$$n = n' + m'. \quad (5)$$

This can be interpreted thus: If n denotes the *position* of a particle, $f(n, m; N)$ is the probability that a particle will be found in the n^{th} position at the N^{th} instant and that it will jump m positions within the

following interval. Hence m may be said to correspond to *velocity*. Equations (4) become

$$f(n, m; N+1) = \sum_{m'} p_{n-m'; m', m; N} f(n-m', m'; N). \quad (6)$$

Using the Kronecker symbol and the "step-up" operators of the calculus of finite differences,

$$E_n = 1 + \Delta_n, \quad E_N = 1 + \Delta_N, \quad (7)$$

these may be written

$$\sum_{m'} (p_{n-m'; m', m; N} E_n^{-m'} - \delta_{mm'} E_N) f(n, m'; N) = 0. \quad (8)$$

If the transition probabilities p are independent of position and time—for a Brownian particle this would mean it is "free" apart from the random impacts from the medium molecules, or it is under a constant external force—that is, if $p_{m', m}$ is the probability of a particle with "velocity" m' acquiring "velocity" m after a single interval, the following formal theory is possible. Equations (8), which become

$$\sum_{m'} (p_{m', m} E_n^{-m'} - \delta_{mm'} E_N) f(n, m'; N) = 0, \quad (9)$$

are a system of linear partial difference equations with constant coefficients in

$$\dots, f(n, m-1; N), f(n, m; N), f(n, m+1; N), \dots \quad (10)$$

As in the theory of systems of linear differential equations with constant coefficients, it follows that each of the $f(n, m; N)$ satisfies the same difference equation, viz., the one constructed with the determinant of the operational coefficients of the system of equations (9),

$$|p_{i,j} E_n^{-i} - \delta_{ij} E_N| f(n, m; N) = 0, \quad (11)$$

or, by a slight transformation,

$$D(p; E_n, E_N) f(n, m; N) = 0, \quad (12)$$

where

$$D(p; E_n, E_N) = |p_{i,j} - \delta_{ij} E_n^i E_N|. \quad (13)$$

Any linear combination of the $f(n, m; N)$ will also satisfy the same difference equation; in particular, for the number density,

$$g(n; N) = \sum_m f(n, m; N), \quad (14)$$

one has

$$D(p; E_n, E_N) g(n; N) = 0. \quad (15)$$

Since in Brownian motion one is mainly interested in the moments \bar{n}_N, \bar{n}_N^2 , etc., it is advantageous to define the moment generating function

$$M(\alpha; N) = \sum_n e^{\alpha n} g(n; N) \quad (16)$$

$$= 1 + \alpha \bar{n}_N + \frac{1}{2} \alpha^2 \bar{n}^2_N + \dots \quad (17)$$

Multiplying (15) by $\exp(\alpha n)$ on the left, expanding the determinant in powers of E_n , passing $\exp(\alpha n)$ through the operator so as to make it a factor of $g(n; N)$ with appropriate adjustment, summing for all n , and remembering that E_n operating on a function not containing n may be replaced by unity, one finally obtains an ordinary difference equation for $M(\alpha; N)$,

$$D(p; e^{-\alpha}, E_N) M(\alpha; N) = 0. \quad (18)$$

In principle, the number density itself can then be found from

$$g(n; N) = \frac{1}{2\pi} \int_0^{2\pi} e^{-in\theta} M(i\theta; N) d\theta. \quad (19)$$

The solution of (18) and the subsequent expansion (17), or the integration (19) are in general very difficult; if, however, only the first two moments are required, it is sufficient to expand the determinant and the moment generating function up to the second power of α only, and to equate coefficients of α and α^2 in (18) to zero. This gives

$$D(p; 1, E_N) \bar{n}_N + \left[\frac{\partial}{\partial \alpha} D(p; e^{-\alpha}, 1) \right] = 0, \quad (20)$$

$$D(p; 1, E_N) \bar{n}^2_N + 2 \left[\frac{\partial}{\partial \alpha} D(p; e^{-\alpha}, E_N) \right]_{\alpha=0} \bar{n}_N + \left[\frac{\partial^2}{\partial \alpha^2} D(p; e^{-\alpha}, 1) \right]_{\alpha=0} = 0. \quad (21)$$

In the special cases where the p 's are symmetrical—no external force—that is, where

$$p_{-m', -m} = p_{m', m} \quad (22)$$

in addition to the necessary relation

$$\sum_m p_{m', m} = 1, \quad (23)$$

one finds that the second terms in (20) and (21) vanish, so that these equations become

$$D(p; 1, E_N) \bar{n}_N = 0, \quad (24)$$

$$D(p; 1, E_N) \bar{n}^2_N = - \left[\frac{\partial^2}{\partial \alpha^2} D(p; e^{-\alpha}, 1) \right]_{\alpha=0}. \quad (25)$$

In the following sections only the simplest applications of cases (22) will be considered.

As noted above, any linear combination of the $f(n, m; N)$ will satisfy the difference equation (12). Thus the mean total energy \mathcal{E}_N and the marginal velocity distribution function $h(m; N)$ at the instant N ,

$$\mathcal{E}_N = \sum_n \sum_m \frac{1}{2} m m^2 f(n, m; N), \quad (26)$$

$$h(m; N) = \sum_n f(n, m; N), \quad (27)$$

are easily seen to satisfy

$$D(p; 1, E_N) \mathcal{E}_N = 0, \quad (28)$$

$$D(p; 1, E_N) h(m; N) = 0. \quad (29)$$

CONSERVATIVE CASE

The simplest case to which the preceding theory may be applied is the one where the only transition probabilities are

$$\left. \begin{array}{cc} p_{-1, -1} & p_{-1, 1} \\ p_{1, -1} & p_{1, 1} \end{array} \right\}, \quad (30)$$

which in virtue of (22) and (23) reduce to

$$\left. \begin{array}{cc} p & 1-p \\ 1-p & p \end{array} \right\}, \quad (31)$$

where p is the probability of persistence of velocity, or motion. There is a persistence proper only if $\frac{1}{2} < p < 1$, while for $p = \frac{1}{2}$ the case reduces to the classical random-walk.

From the definition (13) one finds

$$D(p; e^{-\alpha}, E_N) = E_N^2 - (e^{\alpha} + e^{-\alpha}) p E_N + 2p - 1. \quad (32)$$

Substitution into (24), (25) gives the difference equations

$$(E_N + 1 - 2p)(E_N - 1) \bar{n}_N = 0, \quad (33)$$

$$(E_N + 1 - 2p)(E_N - 1) \bar{n}_N^2 = 2p, \quad (34)$$

of which the solutions are

$$\bar{n}_N = \text{const.} + \text{const.} (2p - 1)^N, \quad (35)$$

$$\bar{n}_N^2 = \frac{p}{1-p} N + \text{const.} (2p - 1)^N + \text{const.} \quad (36)$$

With the special initial condition of a particle starting at instant $N = 0$ from position $n = 0$ with equal probabilities $\frac{1}{2}$ to either direction, one easily calculates

$$\bar{n}_N = 0 \quad (37)$$

$$\bar{n}_N^2 = \frac{p}{1-p} N + \frac{2p-1}{2(1-p)} \{(2p-1)^N - 1\}. \quad (38)$$

The latter formula was first obtained by F  rth (1920), to describe the *apparently* random motion of certain infusoria. Owing to their, if limited, memory these animals will in fact show a certain persistence in motion, and the mean square displacement law is not of the same form as Einstein's formula, (1), of the "approximate," or diffusion theory of Brownian motion. Only with $p = \frac{1}{2}$,—no persistence—follows the formula corresponding to (1), $\bar{n}_N^2 = N$ of the classical random walk.

Now, as F  rth pointed out, the model really does apply to a Brownian particle as well. For, owing to its inertia, a particle moving in a certain direction in a small interval is more likely to continue in this direction in the following interval than to reverse, especially as the interval is made infinitely small. Thus, with the interval tending to zero, and the persistence p tending to unity, (38) gives in the limit the "accurate" mean squared displacement formula (2) for Brownian particles initially already in thermal equilibrium with the medium. (This will show also the divergence from the "approximate" theory in which the interval has to remain finite and sufficiently large for there to be no correlation or persistence of motion in consecutive intervals, hence there $p = \frac{1}{2}$ and the classical random-walk analogy.) Writing

$$x = n\xi, \quad t = N\tau, \quad p = 1 - \frac{1}{2}\gamma\tau, \quad (39)$$

where ξ and τ are the standard distance and time intervals, and γ is a constant—an inverse measure of the persistence—on proceeding to the limit $\xi, \tau \rightarrow 0$, one obtains from (38)

$$\bar{x}^2 = 2 \frac{1}{\gamma} (\xi/\tau)^2 t - 2 \frac{1}{\gamma^2} (\xi/\tau)^2 (1 - e^{-\gamma t}). \quad (40)$$

Now $\frac{1}{2}m(\xi/\tau)^2$ is the standard kinetic energy of a particle in this model; equating it to the equipartition value $\frac{1}{2}kT$, and substituting into (40), one gets (2).

In the present case, (31), the kinetic energy of the motion never changes, and in order to derive by this "persistence" method the formulae for Brownian motion from rest, or arbitrary initial velocity, a less simple model is required. This suggested the generalization in the previous and following sections.

Before leaving this case, it may be noted that here the moment-generating function itself can be evaluated from (18) without difficulty. It is

$$\begin{aligned} M(\alpha; N) = & A_1(\alpha) [p \cosh \alpha + \sqrt{\{(1-p)^2 + p^2 \sinh^2 \alpha\}}]^N \\ & + A_2(\alpha) [p \cosh \alpha - \sqrt{\{(1-p)^2 + p^2 \sinh^2 \alpha\}}]^N, \end{aligned} \quad (41)$$

where $A_1(\alpha)$ and $A_2(\alpha)$ are functions of α depending only on initial conditions. For the special initial conditions taken above (after (36)), which are equivalent to

$$M(\alpha; 0) = 1, \quad M(\alpha; 1) = \cosh \alpha, \quad (42)$$

one finds

$$A_1(\alpha) = \frac{1}{2} \left[1 + \cosh \alpha \left\{ 1 + \left(\frac{p}{1-p} \right)^2 \sinh^2 \alpha \right\}^{-\frac{1}{2}} \right], \quad (43)$$

$$A_2(\alpha) = \frac{1}{2} \left[1 - \cosh \alpha \left\{ 1 + \left(\frac{p}{1-p} \right)^2 \sinh^2 \alpha \right\}^{-\frac{1}{2}} \right], \quad (44)$$

from which (37) and (38) might have been derived. For the case of the classical random-walk, $p = \frac{1}{2}$, and the same initial condition, (41) reduces to

$$M(\alpha; N) = (\cosh \alpha)^N. \quad (45)$$

DISSIPATIVE CASE

The next simple case is the special one in which there is only one standard speed, and a possibility of rest but not of outright reversal. Here the only surviving transition probabilities are

$$\left. \begin{array}{ccc} p_{-1,-1} & p_{-1,0} & \\ p_{0,-1} & p_{0,0} & p_{0,1} \\ & p_{1,0} & p_{1,1} \end{array} \right\}, \quad (46)$$

which because of (22) and (23) can be taken as

$$\left. \begin{array}{ccc} p & 1-p & \\ \frac{1}{2}(1-q) & q & \frac{1}{2}(1-q) \\ & 1-p & p \end{array} \right\}, \quad (47)$$

where p is the probability of persistence in motion, and q is the probability of persistence at rest in the standard time interval. The operational determinant is now

$$\begin{aligned} D(p; e^{-\alpha}, E_N) = & -E_N^3 + (2p \cosh \alpha + q) E_N^2 \\ & + \{(1-p-q-pq) \cosh \alpha - p^2\} E_N \\ & + p(p+q-1). \end{aligned} \quad (48)$$

Hence substitution into (24), (25) gives

$$(E_N - p)(E_N + 1 - p - q)(E_N - 1) \bar{n}_N = 0, \quad (49)$$

$$(E_N - p)(E_N + 1 - p - q)(E_N - 1) \bar{n}_N = (1 - p)(1 - q), \quad (50)$$

with the solutions

$$\bar{n}_N = \text{const.} + \text{const.} (p + q - 1)^N + \text{const.} p^N, \quad (51)$$

$$\begin{aligned} \bar{n}_N^2 = & \frac{(1 + p)(1 - q)}{(1 - p)(2 - p - q)} N + \text{const.} + \text{const.} (p + q - 1)^N \\ & + \text{const.} p^N, \end{aligned} \quad (52)$$

Taking the special initial conditions

$$g(0; 0) = 1 = f(0, -1; 0) + f(0, 0; 0) + f(0, 1; 0), \quad (53)$$

$$\bar{m}_0 = -f(0, -1; 0) + f(0, 1; 0), \quad (54)$$

$$\bar{m}_0^2 = f(0, -1; 0) + f(0, 1; 0), \quad (55)$$

the solutions are found to be

$$\bar{n}_N = \bar{m}_0 \frac{1}{1 - p} (1 - p^N), \quad (56)$$

$$\begin{aligned} \bar{n}_N^2 = & \frac{(1 + p)(1 - q)}{(1 - p)(2 - p - q)} N + \frac{2p + q - 1}{(2 - p - q)(1 - q)} \times \\ & \left\{ \frac{1 - q}{2 - p - q} - \bar{m}_0^2 \right\} \{1 - (p + q - 1)^N\} \\ & - \frac{2p}{(1 - p)(1 - q)} \left\{ \frac{1 - q}{1 - p} - \bar{m}_0^2 \right\} \{1 - p^N\}. \end{aligned} \quad (57)$$

From (29), with the same initial conditions, one finds in particular

$$h(0; N) = h(0; 0) + \left\{ \frac{1 - p}{2 - p - q} - h(0; 0) \right\} \{1 - (p + q - 1)^N\}, \quad (58)$$

so that

$$h(0; \infty) = (1 - p)/(2 - p - q); \quad (59)$$

thus eventually the number of particles in motion and the number of particles at rest will tend, irrespective of initial conditions, to the constant ratio

$$(1 - q)/(1 - p). \quad (60)$$

It can similarly be shown that

$$\bar{m}_N = \bar{m}_0 p^N, \quad (61)$$

consistent with (56); and from (28) that

$$\mathcal{E}_N = \mathcal{E}_0 + (\mathcal{E}_\infty - \mathcal{E}_0) \{1 - (p + q - 1)^N\}. \quad (62)$$

Here

$$\mathcal{E}_\infty = \frac{1}{2} \frac{1-q}{2-p-q} m, \quad (63)$$

from which it can be deduced that

$$\mathcal{E}_{N+1} - \mathcal{E}_N = (2-p-q)(\mathcal{E}_\infty - \mathcal{E}_N). \quad (64)$$

Taking

$$x = n\xi, \quad t = N\tau; \quad p = 1 - \gamma_{10}\tau, \quad q = 1 - \gamma_{01}\tau, \quad (65)$$

where γ_{10} and γ_{01} are clearly transition probabilities per unit time, relations (56) and (61) become in the limit

$$\bar{x} = \bar{u}_0 (1 - e^{-\gamma_{10}t}) / \gamma_{10}, \quad (66)$$

$$\bar{u} = \bar{u}_0 e^{-\gamma_{10}t}, \quad (67)$$

for the dependence on time of the position and velocity of the centre of the distribution. The last equation gives

$$\frac{d}{dt} \bar{u} = -\gamma_{10} \bar{u}, \quad (68)$$

so that the motion of the centre of the distribution is analogous to that of a body moving in a resisting medium, and one can consider γ_{10} as a resistance constance per unit mass.

The ratio of particles "in motion" to those "at rest" is by (60) and (65)

$$\gamma_{01}/\gamma_{10}, \quad (69)$$

and the relation corresponding to (63) is

$$E_\infty = \frac{\gamma_{01}}{\gamma_{01} + \gamma_{10}} \frac{1}{2} m (\xi/\tau)^2. \quad (70)$$

Using this expression, (57) becomes in the limit

$$\begin{aligned} \bar{x}^2 = & 2 \frac{2 E_\infty}{m \gamma_{10}} t + \frac{4 E_\infty}{m (\gamma_{01} + \gamma_{10}) \gamma_{01}} \{1 - e^{-(\gamma_{01} + \gamma_{10})t}\} \\ & - 4 \frac{E_\infty (\gamma_{01} + \gamma_{10})}{m \gamma_{10}^2 \gamma_{01}} (1 - e^{-\gamma_{10}t}) \\ & + \bar{u}_0^2 \left[\frac{2}{\gamma_{10} \gamma_{01}} (1 - e^{-\gamma_{10}t}) - \frac{2}{\gamma_{01} (\gamma_{01} + \gamma_{10})} \{1 - e^{-(\gamma_{01} + \gamma_{10})t}\} \right]. \end{aligned} \quad (71)$$

Thus unlike formula (3) which contains three parameters (kT/m , γ , \bar{u}_0^2), the formula arrived at contains four parameters. It becomes formally identical with (3) if one puts $\gamma_{10} = \gamma_{01} = \gamma$,—number of particles in motion and those at rest tending to equality—and $E_\infty = \frac{1}{2} kT$, the equipartition value. In view of the following interpretation, due to Dr. Fürth, this is, however, not permissible.

Formula (71) is a displacement formula for a quantized motion of two energy states, one of zero energy, and the other degenerate with velocities $\pm \lim (\xi/\tau)$, with γ_{10} and γ_{01} as transition probabilities. The probabilities of these two states in stationary equilibrium at temperature T are determined by the well-known relation from statistical mechanics

$$W(\epsilon_s) = g_s e^{-\epsilon_s/kT}, \quad (72)$$

where g_s is the statistical weight of the energy state ϵ_s . In the present case where ϵ is restricted to $\epsilon_0 = 0$ and ϵ_1 , the corresponding statistical weights are equal since transitions can occur only between ϵ_0 and ϵ_1 , but not between the two (degenerate) states belonging to ϵ_1 . Hence

$$\frac{\gamma_{10}}{\gamma_{01}} = \frac{W(\epsilon_0)}{W(\epsilon_1)} = e^{\epsilon_1/kT}. \quad (73)$$

For the limiting case of very high temperature it clearly follows that

$$\gamma_{10} = \gamma_{01} = \gamma, \quad (\text{say}), \quad (74)$$

or $p = q$. Thus the number of parameters is reduced to three, and it is seen that (71) now becomes formally identical with (3), viz.,

$$\bar{x^2} = 2 \frac{E_\infty}{m\gamma} t + \frac{2E_\infty}{m\gamma^2} (1 - e^{-\gamma t}) - 4 \frac{E_\infty}{m\gamma^2} (1 - e^{-\gamma t}) + \frac{\bar{u}_0^2}{\gamma^2} (1 - e^{-\gamma t}). \quad (75)$$

This is to be expected, as the "classical" results are valid only in the limiting case of very high temperature. However, E_∞ will not be identical with the equipartition value $\frac{1}{2}kT$, since the equipartition theorem holds only for continuous energy distributions whereas the present model is strictly restricted to the energies 0 and ϵ_1 .

From (64), proceeding to the limit, it follows that

$$\frac{d}{dt} E = -(\gamma_{01} + \gamma_{10})(E - E_\infty), \quad (76)$$

$$= -2\gamma(E - E_\infty), \quad (77)$$

so that the model reflects faithfully an interesting property of Brownian motion: the energy follows the usual decay law, or, considering particles as having an absolute kinetic temperature proportional to their random energy, (77) may be interpreted thus: the law according to which the medium supplies energy to the particles is analogous to Newton's law of cooling (*cf.* Prigogine, 1947).

ACKNOWLEDGMENT

The author wishes to thank Dr. R. Fürth, of Birkbeck College, for introducing him to the study of Brownian motion, and for many constructive suggestions which contributed to this paper.

APPENDIX

The following results concerning the classical one-dimensional random-walk may be of interest. The moment-generating function of the distribution in n has been found as, (45),

$$M_{(n)}(\alpha; N) = (\cosh \alpha)^N. \quad (78)$$

Hence, from (19) and (17) follow the well-known distribution itself and the moments

$$g(n; N) = \frac{1}{2^N N} C_{(N-n)/2} \quad (79)$$

$$\overline{n^{2r}}_N = \sum_{+2j+3k+\dots=r} \frac{(2r)! N!}{(N-i-j-k-\dots)! i! j! k! \dots} \times \left(\frac{1}{2!}\right)^i \left(\frac{1}{4!}\right)^j \left(\frac{1}{6!}\right)^k \dots, \quad (80)$$

$$\overline{n^2}_N = N, \quad \overline{n^4}_N = N(3N-2), \quad (81)$$

$$\overline{n^6}_N = N(15N^2 - 30N + 16), \quad (82)$$

$$\overline{n^8}_N = N(105N^3 - 420N^2 + 588N - 272). \quad (83)$$

In view of Einstein's formula (1) it is of some interest to consider the distribution, not of n but of n^2 . From (78) it can be shown that the moment-generating function of this latter distribution is

$$M_{(n^2)}(\alpha; N) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ix^2} M_{(n)}(x\sqrt{2\alpha}; N) dx, \quad (84)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ix^2} (\cosh x\sqrt{2\alpha})^N dx, \quad (85)$$

and hence, or better, directly from (81)-(83),

$$\overline{n^2} = N, \quad \overline{(n^2 - \overline{n^2})^2} = 2N(N-2), \quad (86)$$

$$\overline{(n^2 - \overline{n^2})^3} = 8N(N-1)(N-2), \quad (87)$$

$$\overline{(n^2 - \overline{n^2})^4} = 4N(N-1)(15N^2 - 63N + 68). \quad (88)$$

Hence the discrete probability distribution in n^2 has

$$\overline{n^2} = \text{mean} = N = N, \quad (89)$$

$$\overline{(n^2 - \overline{n^2})^2} = \text{variance} = 2N(N-1) \rightarrow 2N^2, \quad (90)$$

$$\alpha_3 = \frac{(n^2 - \bar{n}^2)^3}{\{(n^2 - \bar{n}^2)^2\}^{3/2}} = \text{skewness} = 2\sqrt{2} \frac{N-2}{\sqrt{N(N-1)}} \rightarrow 2\sqrt{2}, \quad (91)$$

$$\alpha_4 = \frac{(n^2 - \bar{n}^2)^4}{\{(n^2 - \bar{n}^2)^2\}^2} = \text{kurtosis} = \frac{15N^2 - 63N + 68}{N(N-1)} \rightarrow 15. \quad (92)$$

Thus the distribution of n^2 is leptocurtic and of the χ^2 type, or Pearson Type III.

REFERENCES TO LITERATURE

- CHANDRASEKHAR, S., 1943. "Stochastic Problems in Physics and Astronomy", *Rev. Mod. Phys.*, xv, 1-89.
- EINSTEIN, A., 1905. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen", *Ann. der Phys.*, xvii, 549-560.
- FÜRTH, R., 1920. "Die Brownsche Bewegung bei Berücksichtigung einer Persistenz der Bewegungsrichtung. Mit Anwendungen auf die Bewegung lebender Infusorien", *Zeits. f. Phys.*, ii, 244-256.
- KAC, M., 1945. "Random Walk in the Presence of Absorbing Barriers", *Ann. Math. Statistics*, xvi, 62-67.
- , 1947. "Random Walk and the Theory of Brownian Motion", *Amer. Math. Mon.*, liv, 369-391.
- ORNSTEIN, L. S., 1918. "On the Brownian Motion", *Proc. Kon. Ned. Akad. Wet. Amsterdam*, xxi, 96-108.
- PRIGOGINE, I., 1947. *Étude Thermodynamique des Phénomènes irréversibles*, Dunod-Paris.
- RAYLEIGH, Lord, 1919. "On the Problem of Random Vibrations, and of Random Flight in One, Two, or Three Dimensions", *Phil. Mag.*, xxxvii, 321-347.
- , *Scientific Papers*.
- SMOLUCHOWSKI, M. v., 1906. "Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen", *Ann. der Phys.*, xxi, 756-780.
- , 1923. *Ostwalds Klassiker*, No. 207, Leipzig.

(Issued separately, May 30, 1952)

XX.—On the Estimation of Variance and Covariance. By E. H. Lloyd, Imperial College, London. *Communicated by Professor H. LEVY.*

(MS. received September 3, 1951. Revised MS. received November 26, 1951.

Read January 7, 1952)

SYNOPSIS

Suppose we have a number of independent pairs of observations (X_i, Y_i) on two correlated variates (X, Y) , which have constant variances and covariance, and whose expected values are of known linear form, with unknown coefficients: say $\sum p_{ij}a_j$, $\sum q_{ij}b_j$ respectively. The p_{ij} and the q_{ij} are known, the a_j and the b_j are unknown. The paper discusses the estimation of the coefficients, and of the variances and the covariance, and evaluates the sampling variances of the estimates. The argument is entirely free of distributional assumptions.

1. INTRODUCTION

The subject of this paper was suggested by a problem first discussed by Aitken (1948). That problem is here generalized and freed from distributional assumptions.

Suppose we have n independent pairs of observations (X_i, Y_i) , with variances σ_1^2 , σ_2^2 respectively, covariance $\rho\sigma_1\sigma_2$, and expectations which are linear functions of k ($\leq n$) constant coefficients. The problem is to estimate these coefficients, and the variances and the covariance of X and Y , and to evaluate the sampling variances of these estimates.

The expectations have the form

$$\mathcal{E}(X_s) = \sum_{i=1}^k p_{si}a_i, \quad \mathcal{E}(Y_s) = \sum_{i=1}^k q_{si}b_i \quad (s = 1, 2, \dots, n). \quad (1.1)$$

Here the a_i and the b_i are the coefficients to be estimated, and the p_{si} and the q_{si} are any set of numbers determined by the design of the experiment. We assume the matrices $[p_{si}]$, $[q_{si}]$ to be of full rank k .

It will be shown that subject to certain (stated) conditions on the p_{si} and the q_{si} , the best unbiased linear estimates of the a_i and the b_i are unaffected by ignorance of ρ .

When one estimates the a_i and the b_i , unbiased quadratic estimates of σ_1^2 , σ_2^2 and $\rho\sigma_1\sigma_2$ may also be obtained as by-products. Formulae are derived giving the sampling variances of these estimates in terms of the second and fourth central moments of the original observations.

The discussion is conducted entirely in distribution-free terms; in particular, the results do not depend on normality.

2. NOTATION

Variates and observations on variates are denoted by upper case latin letters. Symbols in bold type represent matrices, including vectors, a vector being understood as a *column*-vector. A prime, as usual, denotes transposition. The standard symbol \mathbf{I} represents the unit matrix (this being a slight but unambiguous departure from our rules).

The set of observations X_i are represented by the vector \mathbf{X} , and the set of Y_i 's by \mathbf{Y} . The a_i and the b_i are likewise written as vectors \mathbf{a} , \mathbf{b} . The $(n \times k)$ matrix of the p_{ij} is denoted by \mathbf{p} , and the $(n \times k)$ matrix of the q_{ij} by \mathbf{q} .

We now write (1.1) in matrix form:

$$\mathcal{E}(\mathbf{X}) = \mathbf{p}\mathbf{a}, \quad \mathcal{E}(\mathbf{Y}) = \mathbf{q}\mathbf{b}. \quad (2.1)$$

We assume that the variances σ_1^2 , σ_2^2 of the X_i , Y_i respectively, and their covariance $\rho\sigma_1\sigma_2$, are constant. The variance matrices and the covariance matrix of the observations are then

$$\begin{aligned} U(\mathbf{X}) &= \sigma_1^2 \mathbf{I}, & U(\mathbf{Y}) &= \sigma_2^2 \mathbf{I}, \\ \mathcal{C}(\mathbf{X}, \mathbf{Y}) &= \rho\sigma_1\sigma_2 \mathbf{I}, \end{aligned} \quad (2.2)$$

where

$$\mathcal{C}(\mathbf{X}, \mathbf{Y}) = \mathcal{E}(\mathbf{XY}') - \mathcal{E}(\mathbf{X}) \cdot \mathcal{E}(\mathbf{Y}')$$

and

$$U(\mathbf{X}) = \mathcal{C}(\mathbf{X}, \mathbf{X}).$$

The two matrix equations (2.1) can be combined in the form

$$\mathcal{E}(\mathbf{Z}) = \mathbf{rc} \quad (2.3)$$

where

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} \mathbf{p} & \mathbf{0} \\ \mathbf{0} & \mathbf{q} \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}.$$

The variance matrix of \mathbf{Z} is

$$\mathbf{v} = \begin{pmatrix} \sigma_1^2 \mathbf{I} & \rho\sigma_1\sigma_2 \mathbf{I} \\ \rho\sigma_1\sigma_2 \mathbf{I} & \sigma_2^2 \mathbf{I} \end{pmatrix}. \quad (2.4)$$

3. ESTIMATION OF THE LINEAR COEFFICIENTS

If we consider now the estimation of \mathbf{a} and \mathbf{b} , the Least Squares Theorem gives the best unbiased linear estimate as

$$\hat{\mathbf{c}} = (\mathbf{r}'\mathbf{v}^{-1}\mathbf{r})^{-1}\mathbf{r}'\mathbf{v}^{-1}\mathbf{Z}, \quad (3.1)$$

with variance matrix

$$U(\hat{\mathbf{c}}) = (\mathbf{r}'\mathbf{v}^{-1}\mathbf{r})^{-1}. \quad (3.2)$$

If ρ is unknown, however, an estimate involving \mathbf{v} will not usually be of any practical use. We must then do the best we can, which is to use the two equations (2.1) separately, obtaining the estimates

$$\check{\mathbf{a}} = (\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{X}, \quad \check{\mathbf{b}} = (\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}'\mathbf{Y}. \quad (3.3)$$

We shall refer to this as the "separate" procedure. It will still be worth while to investigate the "joint" procedure which leads to (3.1), however, since it turns out that in certain circumstances the two procedures lead to the same estimates.

3a. THE SPECIAL CASE OF $\mathbf{p} = \mathbf{q}$

Before investigating the general question, we consider the special case examined by Aitken, namely the case where $\mathbf{p} = \mathbf{q}$. We obtain the interesting result that the "separate" estimates here coincide with the "joint" estimates, so that, in this case at least, it appears that we lose nothing by not knowing ρ .

To prove this we develop (3.1), which becomes

$$\hat{\mathbf{c}} = (\mathbf{s}'\mathbf{v}^{-1}\mathbf{s})^{-1}\mathbf{s}'\mathbf{v}^{-1}\mathbf{Z},$$

where

$$\mathbf{s} = \begin{pmatrix} \mathbf{p} & \mathbf{0} \\ \mathbf{0} & \mathbf{p} \end{pmatrix}.$$

The submatrices into which \mathbf{v} , given by (2.4), is partitioned are scalar and thus commutative. Hence we have at once

$$\mathbf{v}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_1^2\mathbf{I} & -\rho\sigma_1\sigma_2\mathbf{I} \\ -\rho\sigma_1\sigma_2\mathbf{I} & \sigma_1^2\mathbf{I} \end{pmatrix}. \quad (3.4)$$

Thus

$$\mathbf{s}'\mathbf{v}^{-1}\mathbf{s} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_1^2\mathbf{p}'\mathbf{p} & -\rho\sigma_1\sigma_2\mathbf{p}'\mathbf{p} \\ -\rho\sigma_1\sigma_2\mathbf{p}'\mathbf{p} & \sigma_1^2\mathbf{p}'\mathbf{p} \end{pmatrix}.$$

The submatrices into which this matrix is partitioned are again commutative. The inverse is thus

$$(\mathbf{s}'\mathbf{v}^{-1}\mathbf{s})^{-1} = \begin{pmatrix} \sigma_1^2(\mathbf{p}'\mathbf{p})^{-1} & \rho\sigma_1\sigma_2(\mathbf{p}'\mathbf{p})^{-1} \\ \rho\sigma_1\sigma_2(\mathbf{p}'\mathbf{p})^{-1} & \sigma_2^2(\mathbf{p}'\mathbf{p})^{-1} \end{pmatrix}, \quad (3.5)$$

whence, finally,

$$\hat{\mathbf{c}} = \begin{pmatrix} (\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{X} \\ (\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{Y} \end{pmatrix}. \quad (3.6)$$

These results are identical with the "separate" estimates (3.3).

3b. THE GENERAL CASE: $\mathbf{p} \neq \mathbf{q}$

We now consider whether results similar to the above can be obtained when \mathbf{p} and \mathbf{q} are no longer equal. The matrix $\mathbf{r}'\mathbf{v}^{-1}\mathbf{r}$ of (3.1) now has the form

$$\sigma_1^2\sigma_2^2(1-\rho^2)\mathbf{r}'\mathbf{v}^{-1}\mathbf{r} = \begin{pmatrix} \sigma_1^2\mathbf{p}'\mathbf{p} & -\rho\sigma_1\sigma_2\mathbf{p}'\mathbf{q} \\ -\rho\sigma_1\sigma_2\mathbf{q}'\mathbf{p} & \sigma_1^2\mathbf{q}'\mathbf{q} \end{pmatrix}. \quad (3.7)$$

The submatrices here being no longer commutative, inversion of the

matrix must be carried out from first principles. We therefore rewrite (3.1) in the form

$$(\mathbf{r}'\mathbf{v}^{-1}\mathbf{r})\hat{\mathbf{c}} = \mathbf{r}'\mathbf{v}^{-1}\mathbf{Z}.$$

Using (3.7), this becomes

$$\begin{aligned}\sigma_1^2\mathbf{p}'\mathbf{p}\hat{\mathbf{a}} - \rho\sigma_1\sigma_2\mathbf{p}'\mathbf{q}\hat{\mathbf{b}} &= \sigma_1^2\mathbf{p}'\mathbf{X} - \rho\sigma_1\sigma_2\mathbf{p}'\mathbf{Y} \\ -\rho\sigma_1\sigma_2\mathbf{q}'\mathbf{p}\hat{\mathbf{a}} + \sigma_2^2\mathbf{q}'\mathbf{q}\hat{\mathbf{b}} &= -\rho\sigma_1\sigma_2\mathbf{q}'\mathbf{X} + \sigma_2^2\mathbf{q}'\mathbf{Y}.\end{aligned}$$

Solving these two matrix equations for $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ we obtain

$$\begin{aligned}[\mathbf{p}'(\mathbf{I} - \rho^2\mathbf{q}(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}')\mathbf{p}]\hat{\mathbf{a}} &= \mathbf{p}'[\mathbf{I} - \rho^2\mathbf{q}(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}']\mathbf{X} \\ &\quad - \rho\left(\frac{\sigma_1}{\sigma_2}\right)\mathbf{p}'[\mathbf{I} - \mathbf{q}(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}']\mathbf{Y}\end{aligned}\quad (3.8a)$$

and

$$\begin{aligned}[\mathbf{q}'(\mathbf{I} - \rho^2\mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}')\mathbf{q}]\hat{\mathbf{b}} &= \mathbf{q}'[\mathbf{I} - \rho^2\mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}']\mathbf{Y} \\ &\quad - \rho\left(\frac{\sigma_2}{\sigma_1}\right)\mathbf{q}'[\mathbf{I} - \mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}']\mathbf{X}.\end{aligned}\quad (3.8b)$$

These results are to be contrasted with the "separate" estimates (3.3). It will be seen that the "separate" estimates will coincide with these "joint" estimates if and only if the coefficient of \mathbf{Y} in (3.8a) and of \mathbf{X} in (3.8b) both vanish; i.e., if

$$\mathbf{p}'[\mathbf{I} - \mathbf{q}(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}'] = \mathbf{q}'[\mathbf{I} - \mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'] = \mathbf{0}. \quad (3.9)$$

These conditions may be written as

$$\mathbf{q}(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}'\mathbf{p} = \mathbf{p}, \quad \mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{q} = \mathbf{q}. \quad (3.10)$$

These two relations are in fact equivalent. Premultiplying the first, for example, by \mathbf{p}' , we obtain

$$\mathbf{p}'\mathbf{q}(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}'\mathbf{p} = \mathbf{p}'\mathbf{p};$$

now the matrix $\mathbf{p}'\mathbf{p}$ is non-singular, and since $\mathbf{p}'\mathbf{q}$ and its transpose $\mathbf{q}'\mathbf{p}$ are square matrices it follows that $\mathbf{p}'\mathbf{q}$ is itself non-singular. Thus

$$(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}'\mathbf{p} = (\mathbf{p}'\mathbf{q})^{-1}\mathbf{p}'\mathbf{p},$$

so that

$$\mathbf{q}'\mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{q} = \mathbf{q}'\mathbf{q}.$$

If we now postmultiply the first relation (3.10) by $(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{q}$ we obtain

$$\begin{aligned}\mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{q} &= \mathbf{q}(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}'\mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{q} \\ &= \mathbf{q}(\mathbf{q}'\mathbf{q})^{-1}\mathbf{q}'\mathbf{q} = \mathbf{q}\end{aligned}$$

by (3.1), and this is the second of the relations* (3.10).

* This proof, which is much better than my original proof, is due to the referees.

The two relations (3.10) are thus equivalent, and either gives a necessary and sufficient condition for the coincidence of the two kinds of estimate.

4. ESTIMATION OF VARIANCE AND COVARIANCE

Aitken has shown that in the case where $p = q$ a bilinear form in the residuals of X and Y furnishes an unbiased estimate of the covariance $\rho\sigma_1\sigma_2$; and has extended this to some special cases of $p \neq q$ where p and q are suitably related to one another. His derivations assume the residuals to be normally distributed.

However, these results can be established without appealing to normality, and without requiring any special relationship between p and q . This is very satisfying in that it preserves the distribution-free nature of Least Squares estimation.

We assume, as before, that

$$\mathcal{E}(X) = pa, \quad \mathcal{E}(Y) = qb,$$

and we estimate a and b by the "separate" procedure (3.3). We then obtain the residual vectors:

$$\begin{aligned} E &= X - p\check{a} = mX, \\ F &= Y - q\check{b} = nY, \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} m &= I - p(p'p)^{-1}p' = m' = m^2, \\ n &= I - q(q'q)^{-1}q' = n' = n^2. \end{aligned} \tag{4.2}$$

We note that

$$mp = nq = 0,$$

whence, by (2.1), E and F have zero expectations. Their variance matrices are then

$$\begin{aligned} U(E) &= \mathcal{E}(EE') = \sigma_1^2 m, \\ U(F) &= \mathcal{E}(FF') = \sigma_2^2 n. \end{aligned} \tag{4.3}$$

4a. ESTIMATION OF THE VARIANCES

We consider first the estimation of σ_1^2 from the squared residuals. The expectations of the individual squared residuals are the diagonal elements of the variance matrix (4.3); hence the expectation of the sum $E'E$ of squared residuals is the sum of these diagonal elements, i.e., the trace of the variance matrix.* Thus

$$\mathcal{E}(E'E) = \sigma_1^2 \text{tr}(m). \tag{4.4}$$

* For this remark, which greatly simplifies my original derivation, I am indebted to the referees.

To evaluate this we note that

$$\text{tr}(\lambda\mu) = \text{tr}(\mu\lambda)$$

for any matrices λ, μ for which the two products $\lambda\mu$ and $\mu\lambda$ both exist. Thus

$$\begin{aligned}\text{tr}\{\mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\} &= \text{tr}\{(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\mathbf{p}\} \\ &= \text{tr}(\mathbf{I}_k) = k,\end{aligned}$$

where \mathbf{I}_k denotes the unit matrix of order k . It follows that

$$\begin{aligned}\text{tr}(\mathbf{m}) &= \text{tr}\{\mathbf{I}_n - \mathbf{p}(\mathbf{p}'\mathbf{p})^{-1}\mathbf{p}'\} \\ &= n - k.\end{aligned}$$

Thus

$$\mathcal{E}(\mathbf{E}'\mathbf{E}) = (n - k)\sigma_1^2,$$

and so

$$\check{\sigma}_1^2 = (\mathbf{E}'\mathbf{E})/(n - k) \quad (4.5)$$

is an unbiased estimate of σ_1^2 . Similarly

$$\check{\sigma}_2^2 = (\mathbf{F}'\mathbf{F})/(n - k)$$

is an unbiased estimate of σ_2^2 .

These results are of course familiar. The present derivation is given however since it suggests a method of estimating the covariance $\rho\sigma_1\sigma_2$.

4b. ESTIMATE OF COVARIANCE

We consider now the products of residuals E_iF_i . The expectation of such a product is a diagonal element of the covariance matrix $\mathcal{C}(\mathbf{E}, \mathbf{F})$. The expectation of the sum $\mathbf{E}'\mathbf{F}$ of products of residuals is therefore the trace of this matrix.*

Now

$$\begin{aligned}\mathcal{C}(\mathbf{E}, \mathbf{F}) &= \mathcal{C}(\mathbf{mX}, \mathbf{nY}) = \mathbf{m} \cdot \mathcal{C}(\mathbf{X}, \mathbf{Y}) \cdot \mathbf{n}' \\ &= \rho\sigma_1\sigma_2\mathbf{mn},\end{aligned}$$

by (2.2). Taking the trace of this we therefore have

$$\mathcal{E}(\mathbf{E}'\mathbf{F}) = \rho\sigma_1\sigma_2 \text{tr}(\mathbf{mn}), \quad (4.6)$$

and our unbiased estimate of the covariance is

$$(\rho\check{\sigma}_1\check{\sigma}_2) = (\mathbf{E}'\mathbf{F})/\text{tr}(\mathbf{mn}). \quad (4.7)$$

In general this last expression cannot be further simplified since there is no simple expansion for the trace of a product. In any specific application, of course, $\text{tr}(\mathbf{mn})$ can easily be computed.

* For this remark, which greatly simplifies my original derivation, I am indebted to the referees.

In the special case where \mathbf{p} is equal to \mathbf{q} , and, more generally, when conditions (3.10) hold, we can however find the trace explicitly, since then

$$\mathbf{m} = \mathbf{n},$$

whence

$$\text{tr}(\mathbf{mn}) = \text{tr}(\mathbf{m}^2) = \text{tr}(\mathbf{m}) = n - k.$$

Under these conditions therefore our estimate is

$$(\rho\check{\sigma}_1\sigma_2) = (\mathbf{E}'\mathbf{F})/(n - k). \quad (4.8)$$

This result was obtained by Aitken, for normally distributed residuals.

5. SAMPLING VARIANCE OF THE ESTIMATED VARIANCES AND COVARIANCE

Aitken has shown that, in the case of normally correlated variates, with $\mathbf{p} = \mathbf{q}$, the estimate

$$(\rho\check{\sigma}_1\sigma_2) = (\mathbf{E}'\mathbf{F})/(n - k)$$

of the covariance has sampling variance given by

$$U(\rho\check{\sigma}_1\sigma_2) = (1 + \rho^2)\sigma_1^2\sigma_2^2/(n - k). \quad (5.1)$$

This may be compared with the sampling variance of $\check{\sigma}^2$ obtained from the sum of squared residuals in the well-known univariate normal case:

$$U(\check{\sigma}^2) = 2\sigma^4/(n - k). \quad (5.2)$$

It is important for the satisfactory completion of the Least-Squares Theorem that both these results (5.1) and (5.2) be derived by distribution-free arguments.

5a. SAMPLING VARIANCE OF THE ESTIMATED VARIANCE

The estimate $\check{\sigma}_1^2$ given by (4.5) may be written

$$(n - k)\check{\sigma}_1^2 = \mathbf{X}'\mathbf{m}\mathbf{X},$$

whence

$$(n - k)^2 U(\check{\sigma}_1^2) = \mathcal{E}(\mathbf{X}'\mathbf{m}\mathbf{X})^2 - (n - k)^2\sigma_1^4. \quad (5.3)$$

To evaluate this expression we put

$$\mathbf{X} = \mathbf{U} + \mathbf{p}\mathbf{a},$$

so that

$$\mathcal{E}(\mathbf{U}) = 0,$$

while the variance matrix and the independence properties of \mathbf{U} are the same as those of \mathbf{X} .

Since $\mathbf{m}\mathbf{p} = 0$, it follows that

$$\mathbf{X}'\mathbf{m}\mathbf{X} = \mathbf{U}'\mathbf{m}\mathbf{U},$$

and in order to evaluate (5.3) we need to find the expectation of $(U'm U)^2$, or

$$(\sum_i m_{ii} U_i^2 + \sum_{i \neq j} m_{ij} U_i U_j)^2.$$

Since the U_i are mutually independent, with zero expectation, the only surviving terms in this expansion will be those in U_i^4 and $U_i^2 U_j^2$.

Let us denote $\mathcal{E}(U_i^4)$ by $\beta_i \sigma_1^2$. We then have

$$\begin{aligned} \mathcal{E}(X'm X)^2 &= \left\{ \sum_i \beta_i m_{ii}^2 + \sum_{i \neq j} (m_{ii} m_{jj} + 2m_{ij}^2) \right\} \sigma_1^4 \\ &= \left\{ \sum_i (\beta_i - 3) m_{ii}^2 + \sum_{i, j} (m_{ii} m_{jj} + 2m_{ij}^2) \right\} \sigma_1^4. \end{aligned} \quad (5.4)$$

We reduce the last term as follows. Firstly we note that

$$\begin{aligned} \sum_{i, j} (m_{ii} m_{jj}) &= (\sum_i m_{ii})^2 = \{\text{tr}(\mathbf{m})\}^2 \\ &= (n - k)^2. \end{aligned}$$

Secondly we have

$$\begin{aligned} \sum_{i, j} m_{ij}^2 &= \sum_i (\sum_j m_{ij} m_{ji}), & \text{since } \mathbf{m} &= \mathbf{m}' \\ &= \sum_i (\mathbf{m}^2)_{ii} = \sum_i m_{ii}, & \text{since } \mathbf{m} &= \mathbf{m}^2 \\ &= n - k. \end{aligned}$$

Equation (5.4) therefore becomes

$$\mathcal{E}(X'm X)^2 = \sigma_1^4 \sum_i (\beta_i - 3) m_{ii}^2 + \{(n - k)^2 + 2(n - k)\} \sigma_1^4.$$

Using this expression in (5.3) we finally obtain the sampling variance of $\check{\sigma}_1^2$ as

$$(n - k)^2 U(\check{\sigma}_1^2) = \sigma_1^4 \sum_i (\beta_i - 3) m_{ii}^2 + 2(n - k) \sigma_1^4. \quad (5.5)$$

(This formula, of course, also applies to univariate Least Squares analysis. It has previously been derived by Hsu (1938).)

In the special case where $\beta_i = 3$ (as in the normal distribution), the last expression reduces to the familiar result

$$U(\check{\sigma}_1^2) = 2\sigma_1^4/(n - k).$$

For this last expression to hold, however, normality is not strictly a necessary condition: the only requirement is that the (independent) observations X_i all have the same variance, and normal kurtosis.

5b. SAMPLING VARIANCE OF THE ESTIMATED COVARIANCE

A very similar argument yields the sampling variance of the estimate of covariance. We have:

$$(n-k)(\rho\check{\sigma}_1\sigma_2) = \mathbf{E}'\mathbf{F} = \mathbf{X}'\mathbf{m}'\mathbf{n}\mathbf{Y} \\ = \mathbf{X}'\mu\mathbf{Y},$$

say, where

$$\mu = \mathbf{m}'\mathbf{n} \ (\neq \mu').$$

Let

$$\mathbf{X} = \mathbf{U} + \mathbf{p}\mathbf{a}, \quad \mathbf{Y} = \mathbf{V} + \mathbf{q}\mathbf{b},$$

so that

$$\mathcal{E}(\mathbf{U}) = \mathcal{E}(\mathbf{V}) = \mathbf{0}.$$

Then

$$\mathbf{E}'\mathbf{F} = \mathbf{U}'\mu\mathbf{V},$$

and

$$(n-k)^2 U(\rho\check{\sigma}_1\sigma_2) = \mathcal{E}(\mathbf{U}'\mu\mathbf{V})^2 - (n-k)^2(\rho\sigma_1\sigma_2)^2. \quad (5.6)$$

In evaluating the expected value of $(\mathbf{U}'\mu\mathbf{V})^2$ we note that the U_i, V_j are independent when $i \neq j$, with zero expectations, so that in squaring the bilinear form we need retain only terms in $U_i^2V_j^2$ and $U_iV_iU_jV_j$. Neglecting terms of zero expectation, we have

$$(\mathbf{U}'\mu\mathbf{V})^2 = (\sum_{i,j} \mu_{ij}U_iV_j)^2 = \sum_i \mu_{ii}^2 + \sum_{i \neq j} \mu_{ij}^2 U_i^2 V_j^2 \\ + \sum_{i \neq j} (\mu_{ij}^2 + \mu_{ii}\mu_{jj})U_iU_jV_iV_j. \quad (5.7)$$

We now take expectations, noting that

$$\mathcal{E}(U_i^2V_j^2) = \sigma_1^2\sigma_2^2, \\ \mathcal{E}(U_iU_jV_iV_j) = \mathcal{E}(U_iV_i) \cdot \mathcal{E}(U_jV_j) = (\rho\sigma_1\sigma_2)^2.$$

Let us denote $\mathcal{E}(U_i^2V_i^2)$ by $v_i\sigma_1^2\sigma_2^2$. Then (5.7) yields:

$$\mathcal{E}(\sum_{i,j} \mu_{ij}U_iV_j)^2 \\ = \sigma_1^2\sigma_2^2 \{ \sum_i v_i \mu_{ii}^2 + \sum_{i \neq j} \mu_{ij}^2 + \rho^2 \sum_{i \neq j} (\mu_{ij}^2 + \mu_{ii}\mu_{jj}) \} \\ = \sigma_1^2\sigma_2^2 [\sum_i \{ v_i - (1 + 2\rho^2) \} \mu_{ii}^2 + \sum_{i,j} \mu_{ij}^2 \\ + \rho^2 \sum_{i,j} (\mu_{ij}^2 + \mu_{ii}\mu_{jj})]. \quad (5.8)$$

This may be somewhat simplified by noting that

$$\sum_{i,j} \mu_{ij}^2 = \sum_i (\mu\mu')_{ii} = \text{tr}(\mu\mu') \\ = \text{tr}(\mathbf{m}'\mathbf{n}\mathbf{m}) = \text{tr}(\mathbf{m}\mathbf{m}'\mathbf{n}) = \text{tr}(\mathbf{m}\mathbf{n}) \\ = \text{tr}(\mu) = h,$$

say, and

$$\begin{aligned}\sum \mu_{ii}\mu_{jj} &= (\sum \mu_{ii})(\sum \mu_{jj}) \\ &= \{\text{tr}(\mu)\}^2 = h^2.\end{aligned}$$

Using these values in (5.6) and (5.8) we obtain

$$\begin{aligned}\left(\frac{n-k}{\sigma_1\sigma_2}\right)^2 U(\rho\check{\sigma}_1\sigma_2) &= \sum_i \{v_i - (1 + 2\rho^2)\} \mu_{ii}^2 + h(1 + \rho^2) \\ &\quad + \rho^2\{h^2 - (n-k)^2\}.\end{aligned}$$

This is the general expression for the sampling variance of the estimate of covariance.

In the special case when $m = n$ (and, in particular, when $p = q$) we have

$$h = n - k,$$

and (5.10) becomes

$$\left(\frac{n-k}{\sigma_1\sigma_2}\right)^2 U(\rho\check{\sigma}_1\sigma_2) = \sum_i \{v_i - (1 + 2\rho^2)\} m_{ii}^2 + (n-k)(1 + \rho^2). \quad (5.9)$$

In the case of normality, $v_i = 1 + 2\rho^2$, and (5.9) reduces to Aitken's result:

$$U(\rho\check{\sigma}_1\sigma_2) = (1 + \rho^2) \sigma_1^2 \sigma_2^2 / (n - k).$$

ACKNOWLEDGMENTS

The author is deeply indebted to the referees for much detailed and constructive advice.

REFERENCES TO LITERATURE

- AITKEN, A. C., 1948. "On a Problem in Correlated Errors", *Proc. Roy. Soc. Edin.*, A, LXII, 273-277.
HSU, P. L., 1938. "On the Best Unbiased Quadratic Estimate of the Variance" *Statist. Res. Memoirs*, II, 90, (73).

(Issued separately May 30, 1952)

XXI.—The Statistical Theory of Stiff Chains. By H. E. Daniels, M.A.(Cantab.), Ph.D.(Edin.), Statistical Laboratory, University of Cambridge. *Communicated by Professor A. C. AITKEN, F.R.S.*

(MS. received July 31, 1951. Read December 3, 1951.

Revised MS. received January 3, 1952.)

SYNOPSIS

The paper is concerned with the distributional properties of Markoff chains in two and three dimensions where the transition probability for the length of a step and its orientation relative to that of the previous step is specified.

The discrete two-dimensional chain of n steps is first discussed, and by the use of moving axes an equation relating characteristic functions of the end-point distribution for successive values of n is obtained. The corresponding differential equation for the limiting chain with continuous first derivatives is given and asymptotic solutions for long chains are found.

The three-dimensional chain is similarly treated in terms of moving axes, and the limiting continuous chain is again discussed. Finally the same methods are applied to the discrete chain of equal steps to obtain the asymptotic form of the end-point distribution for long chains.

1. INTRODUCTION.—Since Rayleigh's solution of the problem of "random flights" (1919), distribution problems associated with freely linked random chains have been widely discussed under the name of "random walk" problems (see, e.g., Bartlett (1949)). In the present paper we are concerned with the statistical behaviour of what will be called *stiff* chains, in which the orientation of a given link of the chain is influenced by those of the neighbouring links. In Kuhn's theory of rubber elasticity, for example, the chain molecules are stiff in this sense, and are replaced for simplicity by "equivalent" chains containing a smaller number of freely jointed links. Another example of a stiff chain is provided by the path of a heavy particle moving through an atmosphere of light particles, the deflection at each collision being relative to the direction of motion prior to the collision. This multiple scattering problem is considered by Rossi and Greisen (1941), and by Moyal (1950) who gives comprehensive references.

The term "chain" is used here in its widest sense, and since the chain is in general regarded as having a specific direction, it is convenient to distinguish one end as the *initial* or *starting point* and to regard the chain as proceeding in a series of *steps* from one point to the next. The purpose of this paper is to develop the theory of stiff chains which have a Markoff property, the transition probability being the chance that a step has a given length and given orientation relative to that of the

previous step. Many physical chain processes are adequately represented by a model of this kind.

If the orientation of the final step is alone of interest, its distribution can be obtained by the methods of Perrin (1928) and Goudsmit and Saunderson (1940), the problem being essentially that of random walk on a sphere. When step length is taken into account the analysis is very complicated, and even in the case considered here of end-separation of the chain it has so far only been possible to obtain limiting approximations to the distribution.

It is assumed throughout that each point of space may be crossed by the chain any number of times.

2. THE TWO-DIMENSIONAL CHAIN.—Let a be the length of one of the steps of the chain, and let α be the angle it makes with the previous step, with the convention that $\alpha = 0$ when the steps are in line. Set up Cartesian axes with origin at the *end-point* of the chain so that the axis of x lies along the final step. There is a translation and rotation of axes with each step, the addition of the final step transforming the co-ordinates of the initial point from (x', y') to (x, y) according to the relation

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \quad (2.1)$$

which incidentally defines the sense of α . The reason for the unusual choice of axes is that the process is expressible as a Markoff chain in (x, y) alone, the relation between successive probabilities being

$$\begin{aligned} f(x, y; n) &= E f(x', y'; n-1) \\ &= \int_0^{2\pi} da \int_{-\pi}^{\pi} dx f(x', y'; n-1) g(a, \alpha; n), \end{aligned} \quad (2.2)$$

where E denotes expectation and $f(x, y; n)$, $g(u, \alpha; n)$ are the frequency functions for the variables concerned. The formulation as a Markoff process in terms of fixed axes would require the introduction of a third variable representing the orientation of the final step.

The characteristic functions corresponding to f and g are

$$\phi(\xi, \eta) = E e^{i\xi x + i\eta y}$$

$$\gamma(u, v) = E e^{iu\alpha + ivx}.$$

Then

$$\phi(\xi, \eta; n) = \int_0^{2\pi} e^{iu\alpha} da \int_{-\pi}^{\pi} \phi(\xi \cos \alpha + \eta \sin \alpha, -\xi \sin \alpha + \eta \cos \alpha; n-1) \cdot g(a, \alpha; n) dx.$$

Writing

$$\phi(\rho \cos \psi, \rho \sin \psi; n) = \Phi(\rho, \psi; n)$$

this reduces to

$$\Phi(\rho, \psi; n) = \int_0^{\infty} e^{i\alpha\rho \cos \psi} d\alpha \int_{-\pi}^{\pi} \Phi(\rho, \psi - \alpha; n-1) g(\alpha, \alpha; n) d\alpha, \quad (2.3)$$

or, in operational form [cf. Bartlett (1)],

$$\Phi(\rho, \psi; n) = \gamma(\rho \cos \psi, i \frac{\partial}{\partial \psi}; n) \Phi(\rho, \psi; n-1). \quad (2.4)$$

3. The polar forms of the characteristic and frequency functions are related in the following way: Putting $f(r \cos \theta, r \sin \theta) = F(r, \theta)$,

$$\Phi(\rho, \psi) = \int_0^{\infty} \int_{-\pi}^{\pi} e^{i\rho r \cos(\theta - \psi)} F(r, \theta) r dr d\theta,$$

and if F and Φ are developed as Fourier series,

$$F(r, \theta) = \sum_{s=-\infty}^{\infty} F_s(r) e^{is\theta}, \quad \Phi(\rho, \psi) = \sum_{s=-\infty}^{\infty} \Phi_s(\rho) e^{is\psi},$$

it follows without difficulty that $F_s(r)$ and $\Phi_s(\rho)$ are Fourier-Bessel transforms,

$$\left. \begin{aligned} \Phi_s(\rho) &= 2\pi i^s \int_0^{\infty} F_s(r) J_s(\rho r) r dr \\ F_s(r) &= \frac{1}{2\pi i^s} \int_0^{\infty} \Phi_s(\rho) J_s(\rho r) \rho d\rho \end{aligned} \right\} \quad (3.1)$$

In particular the radial frequency function is obtainable directly from

$$\Phi_0(\rho) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi(\rho, \psi) d\psi \text{ in the form}$$

$$2\pi r F_0(r) = r \int_0^{\infty} \Phi_0(\rho) J_0(\rho r) \rho d\rho \quad (3.2)$$

and μ_j , the j^{th} moment of r^2 about zero, can be found from the expansion

$$\Phi_0(\rho) = \sum_{j=0}^{\infty} (-)^j \frac{\mu_j \rho^{2j}}{2^{2j} j! j!}. \quad (3.3)$$

4. A few specially simple cases may be noted. The process reduces to that of the classical random walk in one and two dimensions respectively when (i) α has both zero mean and zero variance, and (ii) when

$g(a, \alpha)$ is independent of α . A less trivial case is that in which α has zero variance but mean α_0 , not zero or $\pm \pi$. Then

$$\gamma(u, v; n) = \gamma(u, 0; n) e^{i v \cos u}$$

and

$$\begin{aligned} \Phi(\rho, \psi; n) &= \gamma(\rho \cos \psi, 0; n) \Phi(\rho, \psi - \alpha_0; n-1) \\ &= \prod_{j=0}^{n-1} \gamma(\rho \cos (\psi - j\alpha_0), 0; n-j), \end{aligned}$$

since $\Phi(\rho, \psi; 0) = 1$. The limiting distribution for large n will usually be of circular normal form by the central limit theorem. If γ is independent of n it is sufficient that the variance of a should be finite and non-zero. (If the variance of a is zero the chain always lies on a circle of radius $\frac{1}{2} a \operatorname{cosec} \frac{1}{2} \alpha_0$.) When γ depends on n , an additional condition on the third absolute moment of a ensures the limiting circular normal distribution.

5. LIMITING CONTINUOUS CHAIN.—The main purpose of the paper is to consider cases where the α distribution is non-degenerate. In the remainder of the discussion on the two-dimensional problem we consider the limiting continuous chain consisting of a large number of small steps, each deviating slightly in direction from its predecessor.*

Assume that $g(a, \alpha)$ is independent of n . Equation (2.4) may be formally expanded as

$$\begin{aligned} \Phi(\rho, \psi; n) &= \left\{ 1 + E(a) i \rho \cos \psi - E(\alpha) \frac{\partial}{\partial \psi} - \frac{1}{2} E(a^2) \rho^2 \cos^2 \psi \right. \\ &\quad \left. - E(a\alpha) i \rho \cos \psi \frac{\partial}{\partial \psi} + \frac{1}{2} E(\alpha^2) \frac{\partial^2}{\partial \psi^2} + \dots \right\} \\ &\quad \Phi(\rho, \psi; n-1). \end{aligned} \quad (5.1)$$

The order of magnitude of the coefficients as $n \rightarrow \infty$ will determine the limiting form of (5.1); they are here chosen to give limiting chains with continuous first derivatives. Allow $l = nE(a)$ to remain finite and let

$$\frac{E(\alpha)}{E(a)} \rightarrow \kappa, \quad \frac{E(\alpha^2)}{E(a)} \rightarrow 2\lambda,$$

so that $E(a)$, $E(\alpha)$, $E(\alpha^2)$ are all $O(n^{-1})$. Further, assume that the variance of a is of smaller order than $E(a)$. Then $E(a^2)$ and $E(a\alpha)$ are $o(n^{-1})$ and (5.1) is

$$\begin{aligned} \Phi(\rho, \psi; n) &= \left\{ 1 + E(a) i \rho \cos \psi - E(\alpha) \frac{\partial}{\partial \psi} + \frac{1}{2} E(\alpha^2) \frac{\partial^2}{\partial \psi^2} \right. \\ &\quad \left. + o(n^{-1}) \right\} \Phi(\rho, \psi; n-1), \end{aligned}$$

* Mr. D. G. Kendall has pointed out that the equations obtained here by heuristic limit operations can be derived rigorously by the methods of M. Kac (1949).

whence, writing $E(a) = dt$ in the limit,

$$\frac{\partial \Phi}{\partial t} = \Phi \{ \rho \cos \psi - \kappa \frac{\partial \Phi}{\partial \psi} + \lambda \frac{\partial^2 \Phi}{\partial \psi^2} \}. \quad (5.2)$$

The required solution of this equation must reduce to $\Phi = 1$ when $t = 0$, but it has to satisfy a further important condition. Since $\phi(\xi, \eta)$ is a single valued function of (ξ, η) , Φ must have a period 2π in ψ .

6. SYMMETRICAL CASE.—It is simplest to consider first the symmetrical case when $\kappa = 0$,

$$\frac{\partial \Phi}{\partial t} = \Phi \{ \rho \cos \psi + \lambda \frac{\partial^2 \Phi}{\partial \psi^2} \}. \quad (6.1)$$

The required periodic solution of (6.1) can be written down as a series of Mathieu functions,

$$\Phi = \sum_i C_{2i} ce_{2i} \left(\frac{1}{2}\psi, -\frac{1}{2}\frac{i\rho}{\lambda} \right) e^{-r_{2i}t},$$

the coefficients C_{2i} being determined by orthogonality conditions to make $\Phi = 1$ initially (see McLachlan (3)). But both C_{2i} and the characteristic numbers v_{2i} of ce_{2i} are complicated functions of ρ , and the subsequent inversion of Φ to $F(r, \theta)$ presents formidable difficulties. Approximations can be found for large t and large r (small ρ), but it is best to obtain these by solving the equation directly without appealing to the special properties of Mathieu functions, as the method is of general application.

Introducing the Laplace transform*

$$L = \int_0^\infty \Phi e^{-pt} p dt$$

equation (6.1) becomes

$$(p - i\rho \cos \psi) L - \lambda \frac{\partial^2 L}{\partial \psi^2} = p. \quad (6.2)$$

Since Φ , and hence L , is in this case an even function of ψ , the cosine expansion

$$L = L_0 + 2 \sum_1^\infty L_s \cos s\psi$$

will provide the required periodic solution. The coefficients L_s , which are the transforms of Φ_s (§3), satisfy the equations

$$\left. \begin{aligned} pL_0 - i\rho L_1 &= p \\ -\frac{1}{2}i\rho L_{s-1} + (p + s^2\lambda)L_s - \frac{1}{2}i\rho L_{s+1} &= 0, \quad s \geq 1. \end{aligned} \right\} \quad (6.3)$$

* I prefer the "dimensionless" form incorporating the factor p .

From the first ν of these equations, L_0, L_1, \dots, L_ν are found to be

$$L_0 = \frac{p\Delta_{1,\nu} + 2(\frac{1}{2}i\rho)^{\nu+1} L_{\nu+1}}{2\Delta_{0,\nu} - p\Delta_{1,\nu}} \quad (6.4)$$

$$L_s = \frac{(\frac{1}{2}i\rho)^s p\Delta_{s+1,\nu} + (\frac{1}{2}i\rho)^{s-s+1} (2\Delta_{0,s-1} - p\Delta_{1,s-1}) L_{\nu+1}}{2\Delta_{0,s} - p\Delta_{1,s}} \quad (6.5)$$

($1 \leq s \leq \nu$),

where

$$\Delta_{s,\nu} = \begin{vmatrix} p+s^2\lambda & -\frac{1}{2}i\rho & 0 & 0 & \dots & 0 & 0 & 0 \\ -\frac{1}{2}i\rho & p+(s+1)^2\lambda & -\frac{1}{2}i\rho & 0 & \dots & 0 & 0 & 0 \\ 0 & -\frac{1}{2}i\rho & p+(s+2)^2\lambda & -\frac{1}{2}i\rho & \dots & 0 & 0 & 0 \\ & & & \dots & & & & \\ 0 & 0 & 0 & 0 & \dots & -\frac{1}{2}i\rho & p+(\nu-1)^2\lambda & -\frac{1}{2}i\rho \\ 0 & 0 & 0 & 0 & \dots & 0 & -\frac{1}{2}i\rho & p+\nu^2\lambda \end{vmatrix}$$

and $\Delta_{s,s-1} = 1$, $\Delta_{s,s-k} = 0$, ($k > 1$).

The determinants $\Delta_{s,\nu}$ diverge as $\nu \rightarrow \infty$, but may be replaced by $\Delta_{s,\nu} = (p+s^2\lambda)(p+(s+1)^2\lambda) \dots (p+\nu^2\lambda) \Lambda_{s,\nu}$, where $\Lambda_{s,\nu}$ has unit diagonal elements and converges for all ρ and for all $p \geq 0$. It can then be proved from (6.5) that $L_s \rightarrow 0$ as $s \rightarrow \infty$, and hence that

$$L_0 = \frac{\Lambda_{1,\infty}}{(2\Lambda_{0,\infty} - p\Lambda_{1,\infty})}, \quad L_s = \frac{(\frac{1}{2}i\rho)^s \Lambda_{s+1,\infty}}{(p+\lambda) \dots (p+s^2\lambda)(2\Lambda_{0,\infty} - p\Lambda_{1,\infty})} \quad (s \geq 1).$$

7. ASYMPTOTIC DISTRIBUTION FOR LONG CHAINS.—When t is large one expects intuitively that the distribution will approach the circular normal form with $r^2 = O(t)$ over the effective range of r . It is therefore reasonable to investigate approximations to L_s for small p and ρ with $p = O(\rho^2)$. To indicate the order of the approximations, let T denote a typical value of t . Terms which are $O(p^2)$ or $O(\rho^2)$ are taken to be $O(T^{-\nu})$ in the sense that after the change of scale $t = t'T$, $r = r'T^{\frac{1}{2}}$, $T^{-\nu}$ appears explicitly as a factor in such terms, the corresponding variables p', ρ' being $O(1)$ in T .

Both λ and Λ are $O(1)$, so that $L_s = O(T^{-s/2})$, $s \geq 1$. Using the relation

$$\Delta_{s,\nu} = (p+s^2\lambda)\Delta_{s+1,\nu} + \frac{1}{4}\rho^2\Delta_{s+2,\nu} \quad (7.1)$$

(6.4) and (6.5) become

$$L_0 = \frac{p\Delta_{1,\nu}}{p\Delta_{1,\nu} + \frac{1}{4}\rho^2\Delta_{2,\nu}} + O(T^{-\nu}) \quad (7.2)$$

$$L_0 = \frac{(\frac{1}{2}i\rho)^s p \Delta_{s+1, \nu}}{p \Delta_{1, \nu} + \frac{1}{2}\rho^2 \Delta_{2, \nu}} + O(T^{-s+s/2-1}) \quad (s \geq 1), \quad (7.3)$$

from which approximations to the distribution can be found for large T , as successive terms of an asymptotic expansion. Further application of (7.1) shows (7.2) to be the ν^{th} convergent of the continued fraction

$$L_0 = \frac{p}{p + \frac{\frac{1}{2}\rho^2}{p + \lambda + \frac{\frac{1}{2}\rho^2}{p + 2\lambda + \frac{\frac{1}{2}\rho^2}{p + 3\lambda + \dots}}}} \quad (7.4)$$

all subsequent numerators being $\frac{1}{2}\rho^2$.

The first approximation ($\nu = 1$) is

$$\left. \begin{aligned} L_0 &= \frac{p(p + \lambda)}{p(p + \lambda) + \frac{1}{2}\rho^2} + O(T^{-1}) = \frac{p}{p + \frac{\frac{1}{2}\rho^2}{\lambda}} + O(T^{-1}), \\ L_1 &= \frac{\frac{1}{2}i\rho p}{p(p + \lambda) + \frac{1}{2}\rho^2} + O(T^{-3/2}) = \frac{\frac{1}{2}i\rho}{\lambda} \cdot \frac{p}{\left(p + \frac{\frac{1}{2}\rho^2}{\lambda}\right)} + O(T^{-3/2}), \\ L_s &= O(T^{-s/2}), \quad s \geq 2. \end{aligned} \right\} \quad (7.5)$$

Thus

$$L = \left(1 + \frac{i\rho}{\lambda} \cos \psi\right) \frac{p}{\left(p + \frac{\frac{1}{2}\rho^2}{\lambda}\right)} + O(T^{-1}),$$

which is the Laplace transform of

$$\Phi = \left(1 + \frac{i\rho}{\lambda} \cos \psi\right) e^{-\frac{1}{2}\frac{\rho^2}{\lambda}t} + O(T^{-1}). \quad (7.6)$$

The Cartesian form of (7.6) is

$$\begin{aligned} \phi &= \left(1 + \frac{i\xi}{\lambda}\right) e^{-\frac{1}{2}\frac{t}{\lambda}(\xi^2 + \eta^2)} + O(T^{-1}) \\ &= \exp\left(\frac{i\xi}{\lambda} - \frac{1}{2}\frac{t}{\lambda}(\xi^2 + \eta^2)\right) + O(T^{-1}), \end{aligned} \quad (7.7)$$

which is the characteristic function for a circular normal distribution centred at the point $\left(\frac{1}{\lambda}, 0\right)$, the variance of x or y being t/λ .

If the displacement of the centre is ignored the distribution may be considered circularly normal about the point $(0, 0)$, but the terms neglected are then $O(T^{-1})$ instead of $O(T^{-1/2})$. On the other hand, the distribution of r depends only on L_0 (cf. (3.2)), and is therefore

$$2\pi r F_0(r) = \frac{r\lambda}{t} e^{-\frac{r^2\lambda}{t}} \{1 + O(T^{-1})\},$$

the displacement of the centre having no effect to this order.

For the higher approximation it is best to expand (7.2) and (7.3) as a Laurent series

$$L_s = \sum A_m \left(p + \frac{1}{2} \frac{\rho^2}{\lambda} \right)^m,$$

treating p and ρ^2 as $O(T^{-1})$, the A 's being polynomials in ρ . The inversion of the Laplace transform is

$$\Phi_s = \left(A_1 + A_2 t + A_3 \frac{t^2}{2!} + \dots \right) e^{-\frac{1}{2} \frac{\rho^2}{\lambda} t},$$

values of $m \leq 0$ giving no contribution.* The Fourier components of the frequency function are then most simply obtained from (3.1) using the formula

$$\begin{aligned} \int_0^\infty \rho^{l+2m+1} e^{-\frac{1}{2} \frac{\rho^2}{\lambda} r} J_l(r\rho) d\rho \\ = \frac{2^m r^l}{r^{l+m+1}} \left(\frac{\Gamma(l+m+1)}{\Gamma(l+1)} - \frac{m\Gamma(l+m+1)}{\Gamma(l+2)} \left(\frac{r^2}{2r} \right) \right. \\ \left. + \frac{m(m-1)}{2!} \frac{\Gamma(l+m+1)}{\Gamma(l+3)} \left(\frac{r^2}{2r} \right)^2 + \dots \right) e^{-\frac{r^2}{2r}}, \end{aligned} \quad (7.8)$$

where m is an integer and the series terminates.

Proceeding in this way, we find for $\nu = 2$,

$$\begin{aligned} \Phi_0 &= \left\{ 1 + \frac{1}{2} \frac{\rho^2}{\lambda^2} - \frac{7}{32} \frac{\rho^4}{\lambda^3} t \right\} e^{-\frac{1}{2} \frac{\rho^2}{\lambda} t} + O(T^{-2}) \\ \Phi_1 &= \left\{ \frac{1}{2} \frac{i\rho}{\lambda} + \frac{1}{32} \frac{i\rho^3}{\lambda^3} - \frac{7}{64} \frac{i\rho^5}{\lambda^4} t \right\} e^{-\frac{1}{2} \frac{\rho^2}{\lambda} t} + O(T^{-5/2}) \\ \Phi_2 &= -\frac{1}{16} \frac{\rho^2}{\lambda^2} e^{-\frac{1}{2} \frac{\rho^2}{\lambda} t} + O(T^{-2}), \end{aligned}$$

and the second approximation to the distribution is

$$\begin{aligned} F(r, \theta) r dr d\theta = \frac{\lambda}{t} \frac{r}{2\pi} dr d\theta e^{-\frac{\lambda r^2}{2t}} \left[1 - \frac{3}{4} \cdot \frac{1}{\lambda t} + \frac{5}{4} \frac{r^2}{t^2} - \frac{7}{32} \frac{\lambda r^4}{t^3} \right] \\ + \left[\frac{r}{t} - \frac{3}{2} \frac{r}{\lambda t^2} + \frac{5}{16} \frac{r^3}{t^3} - \frac{7}{32} \frac{\lambda r^5}{t^4} \right] \cos \theta \\ + \frac{1}{8} \frac{r^2}{t^2} \cos 2\theta + O(T^{-2}), \end{aligned} \quad (7.9)$$

which is easily expressible in Cartesian form if required.

* In the inversion of the transforms, the contour of Bromwich's integral is modified where necessary to end in the second and third quadrants.

8. SHORT CONTINUOUS CHAINS.—It is also of interest to examine the behaviour of the distribution for small t , the chain being then almost straight. For this purpose Cartesian co-ordinates are more convenient, and the differential equation (6.1) is replaced by

$$\frac{\partial \phi}{\partial t} = \phi i \xi + \lambda \left\{ \eta^2 \frac{\partial^2 \phi}{\partial \xi^2} - 2\xi \eta \frac{\partial^2 \phi}{\partial \xi \partial \eta} + \xi^2 \frac{\partial^2 \phi}{\partial \eta^2} - \xi \frac{\partial \phi}{\partial \xi} - \eta \frac{\partial \phi}{\partial \eta} \right\}. \quad (8.1)$$

On substituting

$$\phi = \sum \sum \mu_{jk} \frac{(i\xi)^j (i\eta)^k}{j! k!}$$

in (8.1), differential equations are obtained relating the bivariate moments μ_{jk} about the origin. Thus

$$\begin{aligned} \frac{\partial \mu_{10}}{\partial t} &= 1 - \lambda \mu_{10}, & \frac{\partial \mu_{01}}{\partial t} &= -\lambda \mu_{01} \\ \frac{1}{2} \frac{\partial \mu_{20}}{\partial t} &= \mu_{10} + \lambda(\mu_{20} + \mu_{02}), & \frac{\partial \mu_{11}}{\partial t} &= \mu_{01} + 4\lambda \mu_{11} \\ \frac{1}{2} \frac{\partial \mu_{02}}{\partial t} &= \lambda(\mu_{20} + \mu_{02}). \end{aligned}$$

All the moments are initially zero except $\mu_{00} = 1$, and the solutions of these equations are

$$\left. \begin{aligned} \mu_{10} &= \frac{1}{\lambda} (1 - e^{-\lambda t}), & \mu_{01} &= 0 \\ \mu_{20} &= \frac{t}{\lambda} - \frac{3}{4\lambda^2} + \frac{2}{3\lambda^2} e^{-\lambda t} + \frac{1}{12\lambda^2} e^{-4\lambda t} \\ \mu_{02} &= \frac{t}{\lambda} - \frac{5}{4\lambda^2} + \frac{4}{3\lambda^2} e^{-\lambda t} - \frac{1}{12\lambda^2} e^{-4\lambda t} \\ \mu_{11} &= 0. \end{aligned} \right\} \quad (8.2)$$

When t is small we find

$$\mu_{10} \sim t, \quad \text{var } x \sim \frac{1}{3} \lambda^2 t^4, \quad \text{var } y \sim \frac{2}{3} \lambda t^3, \quad (8.3)$$

where $\text{var } x = \mu_{20} - \mu_{10}^2$, $\text{var } y = \mu_{02}$. As might be expected, $\text{var } x$ is of smaller order of magnitude than $\text{var } y$, but an unusual feature is that $\text{var } y$ increases as the cube of the chain length (cf. Rossi and Greisen, (1941), p. 268, Moyal, 1950, p. 1062). This differs essentially from the corresponding result when the deviation y of the end-point is the cumulative sum of n independent random lateral displacements, since $\text{var } y$ is then $O(t)$. The results (8.3) are easily verified directly; for example, suppose all steps to be of equal length a , and let $\alpha_2, \alpha_3 \dots \alpha_n$ be the angles made by each step with the previous one, so that

$$y = a \sum_{j=2}^n \sin(\alpha_2 + \alpha_3 + \dots + \alpha_j) \sim a \sum_{j=2}^n (n-j+1) \alpha_j.$$

Then

$$\sigma_y^2 \sim a^2 \sigma_\alpha^2 \sum_{j=2}^n (n-j+1)^2 \sim \frac{1}{3} n^3 a^2 \sigma_\alpha^2$$

and

$$\lambda = \frac{1}{2} \sigma_\alpha^2 / a, \quad t = na.$$

To examine the limiting form of the distribution for small t , let T be again a typical value of t , now a small quantity. Then if $t = t'T$, (8.3) shows it to be appropriate to write $x = x'T$, $y = y'T^{3/2}$, $\xi = \xi'T^{-1}$, $\eta = \eta'T^{-3/2}$, when (8.1) becomes

$$\frac{\partial \phi}{\partial t'} = \phi i \xi' + \lambda \eta'^2 \frac{\partial^2 \phi}{\partial \xi'^2} + O(T). \quad (8.4)$$

Ignoring $O(T)$, the required solution is

$$\phi = e^{i\xi' t' - \frac{1}{2} \lambda t'^2 \eta'^2},$$

showing that, to the order considered, x has the constant value t and the y distribution is normal about zero with variance $\frac{2}{3} \lambda t^3$ (as may also be deduced from the central limit theorem).

It will be observed that the moment formulae (8.2) are exact. Similar exact formulae for the moments μ_j of r^2 about the origin are more simply obtained from (3.3); if L_ν is expanded as a power series in $\frac{1}{2} i \rho$ and substituted in (6.3), the ensuing recurrence relations between the coefficients $L_{\nu, m}$ enable $L_{0, 2j}$ and hence μ_j to be calculated in closed form if required.

9. GENERAL CASE.—The preceding methods are applicable to the more general equation (5.2),

$$\frac{\partial \Phi}{\partial t} = \Phi i \rho \cos \psi - \kappa \frac{\partial \Phi}{\partial \psi} + \lambda \frac{\partial^2 \Phi}{\partial \psi^2},$$

and it will suffice to indicate the limiting form of the distribution for large t . Again Φ must have period 2π in ψ , but is no longer symmetrical in ψ . Corresponding to (6.2) we obtain

$$(p - i \rho \cos \psi) L + \kappa \frac{\partial L}{\partial \psi} - \lambda \frac{\partial^2 L}{\partial \psi^2} = p, \quad (9.1)$$

where now

$$L = \sum_{s=-\infty}^{\infty} L_s e^{i s \psi}.$$

The coefficients satisfy the equations

$$-\frac{1}{2} i \rho L_{s-1} + (p + i s \kappa + s^2 \lambda) L_s - \frac{1}{2} i \rho L_{s+1} = p \quad (s=0) \\ = 0 \quad (s \neq 0). \quad (9.2)$$

For large T it can be shown as before that $L_s = O(T^{-\frac{1}{2}|s|})$ for $s \geq 1$, and the dominant term of the asymptotic expansion is got by putting $L_s = 0$ when $|s| \geq 2$ and ignoring terms which are $O(T^{-1})$ in the expressions for L_0, L_{+1} . Inverting the Laplace transform we find

$$\Phi_0 = e^{-\frac{1}{2} \frac{\lambda^2 t}{(\lambda^2 + \kappa^2)}} + O(T^{-1})$$

$$\Phi_{\pm 1} = \frac{1}{2} i \rho \frac{(\lambda \mp i \kappa)}{(\lambda^2 + \kappa^2)} e^{-\frac{1}{2} \frac{\lambda^2 t}{(\lambda^2 + \kappa^2)}} + O(T^{-3/2}),$$

so that

$$\Phi = e^{-\frac{1}{2} \frac{\lambda^2 t}{(\lambda^2 + \kappa^2)}} \left\{ 1 + \frac{i \rho}{(\lambda^2 + \kappa^2)} (\lambda \cos \psi + \kappa \sin \psi) \right\} + O(T^{-1}).$$

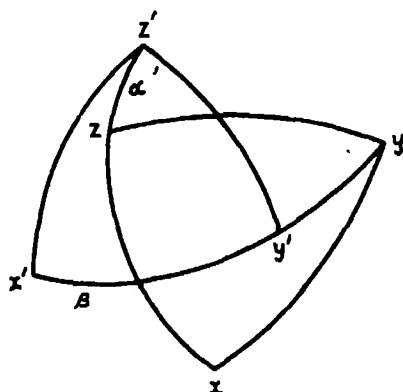


FIG. 1

To the same order, the limiting distribution is therefore circularly normal about the point $\left(\frac{\lambda}{\lambda^2 + \kappa^2}, \frac{\kappa}{\lambda^2 + \kappa^2} \right)$, with variance of x or y equal to $\lambda t / (\lambda^2 + \kappa^2)$.

10. THE THREE-DIMENSIONAL CHAIN.—As before, a system of moving axes is employed with origin at the end-point of the chain. The vector increment for each step is now specified by its length a and two angles α, β defined as follows: To preserve the conventional polar notation the axes are chosen so that the z axis contains the final step, while the (z, x) plane also contains the previous step, that is, the previous z axis. The angle between the last two steps is α as before, while β is the angle turned through by the (z, x) plane. Thus (α, β) are the usual colatitude and azimuth angles specifying the direction of the final step relative to the previous axes. The spherical diagram (fig. 1) shows the angular disposition of the systems of axes (x', y', z') , (x, y, z) adopted respectively before and after a step has been taken.

The co-ordinates of successive points of the chain are connected by the relation

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ a \end{bmatrix} + \begin{bmatrix} \cos \alpha \cos \beta & \cos \alpha \sin \beta & -\sin \alpha \\ -\sin \beta & \cos \beta & 0 \\ \sin \alpha \cos \beta & \sin \alpha \sin \beta & \cos \alpha \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}, \quad (10.1)$$

and the basic probability equation for the chain is

$$f(x, y, z; n) = Ef(x', y', z'; n-1) \\ = \int_0^\infty da \int_0^\pi d\alpha \int_0^{2\pi} d\beta g(a, \alpha, \beta; n) f(x', y', z'; n-1), \quad (10.2)$$

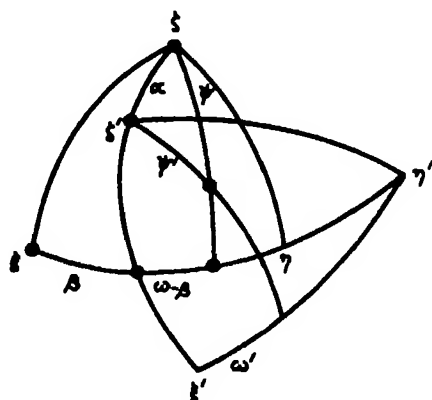


FIG. 2

where $g(a, \alpha, \beta; n)$ is the frequency function for a, α, β at the n^{th} step. Note that for the classical random walk case where all directions of the step are equally likely, g is not independent of the angles but contains the factor $\sin \alpha$.

The characteristic function $\phi(\xi, \eta, \zeta; n) = E e^{i(\xi x + \eta y + \zeta z)}$ satisfies the equation

$$\phi(\xi, \eta, \zeta; n) = E e^{i a \zeta} \phi(\xi', \eta', \zeta'; n-1) \quad (10.3)$$

where

$$\begin{bmatrix} \xi & \eta & \zeta \end{bmatrix} = \begin{bmatrix} \xi' & \eta' & \zeta' \end{bmatrix} \begin{bmatrix} \cos \alpha \cos \beta & \cos \alpha \sin \beta & -\sin \alpha \\ -\sin \beta & \cos \beta & 0 \\ \sin \alpha \cos \beta & \sin \alpha \sin \beta & \cos \alpha \end{bmatrix},$$

and E averages over the a, α, β distribution.

It is again convenient to work in polar co-ordinates

$$x = r \sin \theta \cos \chi \quad \xi = \rho \sin \psi \cos \omega$$

$$\begin{aligned}y &= r \sin \theta \sin \chi & \eta &= \rho \sin \psi \sin \omega \\z &= r \cos \theta & \zeta &= \rho \cos \psi.\end{aligned}$$

Fig. 2 shows the relation between the axes (ξ, η, ζ) and (ξ', η', ζ') , and the corresponding polar angles.

Equation (10.3) becomes

$$\Phi(\rho, \psi, \omega; n) = E e^{i a \rho \cos \psi} \Phi(\rho, \psi', \omega'; n-1), \quad (10.4)$$

where

$$\begin{aligned}\cos \psi' &= \cos \psi \cos \alpha + \sin \psi \sin \alpha \cos (\omega - \beta) \\ \sin \psi' \sin \omega' &= \sin \psi \sin (\omega - \beta).\end{aligned}$$

There does not appear to be a simple operational form similar to (2.4).

For simplicity the discussion is now confined to cases where the angular distribution for each step is axially symmetrical about the direction of the previous step, and is independent of n . The treatment of the general case, though analytically complicated, introduces no new difficulty of principle. Since Φ is assumed independent of ω , we may write

$$\Phi(\rho, \psi, \omega; n) = \frac{1}{2\pi} \Phi(\rho, \psi; n)$$

$$g(a, \alpha, \beta) = \frac{1}{2\pi} g(a, \alpha),$$

and (10.4) reduces to*

$$\Phi(\rho, \psi; n) = \int_0^\infty e^{i a \rho \cos \psi} da \int_0^\pi g(a, \alpha) d\alpha \int_0^{2\pi} \frac{d\omega}{2\pi} \Phi(\rho, \psi'; n-1), \quad (10.5)$$

where

$$\cos \psi' = \cos \psi \cos \alpha + \sin \psi \sin \alpha \cos \omega. \quad (10.6)$$

11. The polar forms of the characteristic and frequency functions are related by formulae analogous to those for two-dimensions. In the axially symmetrical case, writing

$$f(x, y, z) = F(r, \theta, \chi) = \frac{1}{2\pi} F(r, \theta)$$

we find

$$\Phi(\rho, \psi) = \int_0^\infty r^2 dr \int_0^\pi \sin \theta d\theta \int_0^{2\pi} \frac{d\chi}{2\pi} F(r, \theta) e^{i r \rho (\cos \theta \cos \psi + \sin \theta \sin \psi \cos (\chi - \omega))}. \quad (11.1)$$

Series of Legendre polynomials now take the place of Fourier cosine series, and use is made of the expansions

$$e^{i r \rho u} = \sqrt{\frac{2\pi}{r\rho}} \sum_{s=0}^\infty i^s (s + \frac{1}{2}) J_{s+\frac{1}{2}}(r\rho) P_s(u). \quad (11.2)$$

* We now replace $\omega - \beta$ by ω .

$$\begin{aligned}
 P_s(\cos \theta \cos \psi + \sin \theta \sin \psi \cos (\chi - \omega)) \\
 = P_s(\cos \theta) P_s(\cos \psi) + \\
 2 \sum_{m=1}^s \frac{(s-m)!}{(s+m)!} P_s^m(\cos \theta) P_s^m(\cos \psi) \cos m(\chi - \omega), \quad (11.3)
 \end{aligned}$$

where $P_s(u)$ and $P_s^m(u)$ are the ordinary and associated Legendre polynomials. On substituting (11.2) for the exponential in (11.1) and expanding the Legendre polynomials by (11.3), terms in $\cos m(\chi - \omega)$ vanish on integration and (11.1) becomes

$$\begin{aligned}
 \Phi(\rho, \psi) = \sqrt{\frac{2\pi}{\rho}} \sum_{s=0}^{\infty} i^s (s + \frac{1}{2}) P_s(\cos \psi) \\
 \int_0^{\infty} r^3 J_{s+\frac{1}{2}}(r\rho) dr \int_0^{\pi} P_s(\cos \theta) F(r, \theta) \sin \theta d\theta,
 \end{aligned}$$

showing that if

$$\Phi(\rho, \psi) = \sum_{s=0}^{\infty} \Phi_s(\rho) P_s(\cos \psi), \quad F(r, \theta) = \sum_{s=0}^{\infty} F_s(r) P_s(\cos \theta)$$

the coefficients are Fourier-Bessel transforms,

$$\left. \begin{aligned}
 \Phi_s(\rho) &= \sqrt{\frac{2\pi}{\rho}} i^s \int_0^{\infty} F_s(r) J_{s+\frac{1}{2}}(r\rho) r^3 dr \\
 F_s(r) &= \frac{1}{\sqrt{2\pi r}} (-i)^s \int_0^{\infty} \Phi_s(\rho) J_{s+\frac{1}{2}}(r\rho) \rho^3 d\rho.
 \end{aligned} \right\} \quad (11.4)$$

In particular

$$\Phi_0(\rho) = \frac{2}{\rho} \int_0^{\infty} F_0(r) \sin r\rho \cdot r dr,$$

or, since the radial frequency function is $2r^2 F_0(r)$,

$$\Phi_0(\rho) = \sum_{j=0}^{\infty} (-)^j \frac{\mu_j \rho^{2j}}{(2j+1)!}, \quad (11.5)$$

where μ_j is the j^{th} moment of r^2 about the origin.

12. LIMITING CONTINUOUS CHAIN.—The limiting continuous chain is obtained by making a and α small in such a way that $E(\alpha^2)/E(a) \rightarrow 4\lambda$, and assuming $\text{var } a$ to be of smaller order than $E(a)$. The angle α as now defined lies between 0 and π , so that $E(\alpha)$ is always positive except in a degenerate case, whereas in the two-dimensional case α lay between $-\pi$ and π , and the analogue of axial symmetry then had $E(\alpha) = 0$. This is a matter of convention and should cause no confusion.

Expanding the required order,

$$\cos \psi' = \cos \psi + \alpha \sin \psi \cos \omega - \frac{1}{2} \alpha^2 \cos^2 \psi + o(\alpha^2)$$

$$\begin{aligned} \Phi_{n-1}(\rho, \psi') &= \Phi_{n-1}(\rho, \psi) + (\alpha \sin \psi \cos \omega - \frac{1}{2} \alpha^2 \cos^2 \psi) \frac{\partial \Phi_{n-1}}{\partial (\cos \psi)} \\ &\quad + \frac{1}{2} \alpha^2 \sin^2 \psi \cos^2 \omega \frac{\partial^2 \Phi_{n-1}}{\partial (\cos \psi)^2} + o(\alpha^2). \end{aligned}$$

Putting $t = nE(a)$, $dt = E(a)$ and $u = \cos \psi$, (10.5) becomes in the limit

$$\frac{\partial \Phi}{\partial t} = i\rho u \Phi + \lambda \frac{\partial}{\partial u} \left\{ (1-u^2) \frac{\partial \Phi}{\partial u} \right\}. \quad (12.1)$$

The Laplace transform L of Φ with respect to t satisfies the equation

$$(p - i\rho u) L - \lambda \frac{\partial}{\partial u} \left\{ (1-u^2) \frac{\partial L}{\partial u} \right\} = p. \quad (12.2)$$

The analysis then proceeds as in the two-dimensional problem, except that L is expanded as a series of Legendre polynomials in $u = \cos \psi$,

$$L = \sum_{s=0}^{\infty} L_s P_s(u),$$

which, after substitution in (12.2) and simplification by well-known recurrence formulae, leads to the equations

$$\left. \begin{aligned} p L_0 - \frac{1}{3} i\rho L_1 &= p \\ -i\rho \frac{s}{(2s-1)} L_{s-1} + \{p + s(s+1)\lambda\} L_s - \frac{i\rho(s+1)}{(2s+3)} L_{s+1} &= 0. \end{aligned} \right\} \quad (12.3)$$

These are solved as before, the approximation for large T being

$$L_0 = \frac{p \Delta_{1,v}}{\Delta_{0,v}} + O(T^{-v}) = \frac{p \Delta_{1,v}}{p \Delta_{1,v} + \frac{1}{3} \rho^2 \Delta_{2,v}} + O(T^{-1}) \quad (12.4)$$

$$L_s = (2i\rho)^s \frac{(s!)^2}{(2s)!} \frac{p \Delta_{s+1,v}}{(p \Delta_{1,v} + \frac{1}{3} \rho^2 \Delta_{2,v})} + O(T^{-v+s-1}), \quad s \geq 1, \quad (12.5)$$

where

$$\Delta_{s,v} = \begin{vmatrix} p+s(s+1)\lambda & \frac{-i\rho(s+1)}{(2s+3)} & 0 & \dots & 0 & 0 \\ \frac{-i\rho(s+1)}{(2s+1)} & p+(s+1)(s+2)\lambda & \frac{-i\rho(s+2)}{(2s+5)} & \dots & 0 & 0 \\ 0 & \frac{-i\rho(s+2)}{(2s+3)} & p+(s+2)(s+3)\lambda & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{-i\rho v}{(2v+1)} & p+v(v+1)\lambda \end{vmatrix}.$$

If desired, L_0 can be expressed as the continued fraction

$$L_0 = \frac{p}{p + \frac{1}{p + \frac{1 \cdot 3}{p + \frac{2^2}{p + \frac{3^2}{p + \frac{4^2}{p + \dots}}}}}} \rho^2$$

of which (12.4) is the ν^{th} convergent.

The first approximation ($\nu = 1$) is

$$L_0 = \frac{p}{\left(p + \frac{1}{\lambda} \rho^2\right)} + O(T^{-1}), \quad L_1 = \frac{1}{2} \frac{i\rho}{\lambda} \frac{p}{\left(p + \frac{1}{\lambda} \rho^2\right)} + O(T^{-1})$$

$$L_s = O(T^{-1s}), \quad s \geq 2,$$

which transforms back to

$$\Phi = \left(1 + \frac{1}{2} \frac{i\rho}{\lambda} \cos \psi\right) e^{-\frac{1}{2} \frac{\rho^2}{\lambda} t} + O(T^{-1})$$

or, to the same order,

$$\phi = \exp \left\{ \frac{1}{2} \frac{i\rho}{\lambda} - \frac{1}{2} \frac{t}{\lambda} (\xi^2 + \eta^2 + \zeta^2) \right\} + O(T^{-1}).$$

The limiting distribution is of spherical normal form centred at $\left(0, 0, \frac{1}{2\lambda}\right)$, with variance $t/3\lambda$ in each of the three co-ordinates, the radial distribution being unaffected by the eccentricity to the same order.

The second approximation ($\nu = 2$) is

$$\Phi_0 = \left\{ 1 + \frac{1}{12} \frac{\rho^2}{\lambda^2} - \frac{1}{1080} \frac{\rho^4}{\lambda^3} t \right\} e^{-\frac{1}{2} \frac{\rho^2}{\lambda} t} + O(T^{-2})$$

$$\Phi_1 = \left\{ \frac{1}{2} \frac{i\rho}{\lambda} + \frac{1}{180} \frac{i\rho^3}{\lambda^3} - \frac{1}{2160} \frac{i\rho^5}{\lambda^4} t \right\} e^{-\frac{1}{2} \frac{\rho^2}{\lambda} t} + O(T^{-1})$$

$$\Phi_2 = -\frac{1}{18} \frac{\rho^2}{\lambda^2} e^{-\frac{1}{2} \frac{\rho^2}{\lambda} t} + O(T^{-2}),$$

and using (11.4) in conjunction with (7.6), we find

$$\begin{aligned} F(r, \theta) r^2 \sin \theta dr d\theta &= \left(\frac{3\lambda}{t} \right)^{1/2} \frac{r^2 \sin \theta dr d\theta}{\sqrt{2\pi}} \left\{ \left[1 - \frac{3}{8} \cdot \frac{1}{\lambda t} + 2 \frac{r^2}{t^2} - \frac{33}{40} \frac{\lambda r^4}{t^3} \right] \right. \\ &\quad + \left[\frac{3}{8} \frac{r}{t} - \frac{25}{18} \frac{r}{\lambda t^2} + \frac{153}{40} \frac{r^3}{t^3} - \frac{99}{80} \frac{\lambda r^5}{t^4} \right] P_1(\cos \theta) \\ &\quad \left. + \frac{1}{2} \frac{r^2}{t^2} P_2(\cos \theta) + O(T^{-2}) \right\}. \end{aligned} \quad (12.6)$$

The behaviour of the chain for small T can be examined as in §8, using the analogue of (8.1) in cylindrical polar co-ordinates. It is found that, ignoring $O(T)$, z has the constant value t , while (x, y) has a circular normal distribution about zero with variance $\frac{2}{3}\lambda t^2$ in x or y .

13. THE DISCRETE CHAIN OF EQUAL STEPS.—We return to the discrete chain and discuss the important special case where the steps are all of equal size, the modification necessary for variable steps being obvious. The assumption of axial symmetry is retained.

It is convenient to assume the steps to be of unit length, the restriction being removed at the end by writing r/a for r . From (10.5), the generating function

$$G(\rho, \psi; Z) = \sum_{n=0}^{\infty} \Phi(\rho, \psi; n) Z^n$$

is seen to satisfy the equation

$$G(\rho, \psi; Z) = 1 + Z e^{i\rho \cos \psi} \int_0^{\pi} d\alpha \int_0^{2\pi} \frac{d\omega}{2\pi} g(\alpha) G(\rho, \psi'; Z), \quad (13.1)$$

where

$$\cos \psi' = \cos \psi \cos \alpha + \sin \psi \sin \alpha \cos \omega$$

and $g(\alpha)$ is written for $g(\alpha, \alpha)$. Expand G in the form

$$G(\rho, \psi; Z) = \sum_{s=0}^{\infty} G_s(\rho; Z) P_s(\cos \psi).$$

The polynomials $P_s(\cos \psi')$ occurring in the integrand may be effectively replaced by $P_s(\cos \psi) P_s(\cos \alpha)$ using (11.3), and (13.1) becomes

$$\sum_{s=0}^{\infty} G_s P_s(\cos \psi) = 1 + Z e^{i\rho \cos \psi} \sum_{s=0}^{\infty} G_s g_s P_s(\cos \psi), \quad (13.2)$$

where

$$g_s = \int_0^{\pi} g(\alpha) P_s(\cos \alpha) d\alpha, \quad (13.3)$$

so that

$$g(\alpha) = \sin \alpha \sum_{s=0}^{\infty} (s + \frac{1}{2}) g_s P_s(\cos \alpha). \quad (13.4)$$

It follows that

$$G_m = \delta_{m0} + (m + \frac{1}{2}) Z \sum_{s=0}^{\infty} G_s g_s c_{ms}, \quad (13.5)$$

where $\delta_{m0} = 1, 0$ for $m = 0, m > 0$, and

$$\begin{aligned} c_{ms} &= c_{sm} = \int_{-1}^1 e^{i\rho u} P_m(u) P_s(u) du \\ &= 2 P_m \left(\frac{\partial}{i\partial \rho} \right) P_s \left(\frac{\partial}{i\partial \rho} \right) \frac{\sin \rho}{\rho}. \end{aligned} \quad (13.6)$$

In particular,

$$c_{0s} = i^s \sqrt{\frac{2\pi}{\rho}} J_{s+1/2}(\rho).$$

The c 's satisfy the recurrence formula

$$\left. \begin{aligned} c_{m,s+1} &= -i \frac{(2s+1)}{(s+1)} \frac{\partial}{\partial \rho} c_{ms} - \frac{s}{(s+1)} c_{m,s-1} \\ c_{m,-1} &= 0, \end{aligned} \right\} \quad (13.7)$$

from which power series in ρ can be constructed. The constant term in c_{ms} is obviously $2/(2s+1)$. As far as $s=2$, $m=2$, we find

$$\left. \begin{aligned} c_{00} &= 2 - \frac{1}{3} \rho^2 + \frac{1}{60} \rho^4 - \dots \\ c_{01} = c_{10} &= \frac{2}{3} i\rho - \frac{1}{15} i\rho^3 + \frac{1}{420} i\rho^5 - \dots \\ c_{11} &= \frac{2}{3} - \frac{1}{6} \rho^2 + \frac{1}{84} \rho^4 - \dots \\ c_{02} = c_{20} &= -\frac{2}{15} \rho^2 + \frac{1}{105} \rho^4 - \dots \\ c_{12} = c_{21} &= \frac{4}{15} i\rho - \frac{4}{105} i\rho^3 + \frac{1}{315} i\rho^5 - \dots \\ c_{22} &= \frac{2}{5} - \frac{1}{105} \rho^2 + \frac{1}{420} \rho^4 - \dots \end{aligned} \right\} \quad (13.8)$$

14. Consider first the solution of equations (13.5) for the classical random walk where each step has a random direction, that is, $g(\alpha) = \frac{1}{2} \sin \alpha$, $g_0 = 1$, $g_s = 0$ ($s \geq 1$). In this case

$$G_m = \delta_{m0} + (m + \frac{1}{2}) Z G_0 c_{m0},$$

so that

$$G_0 = (1 - \frac{1}{2} Z c_{00})^{-1} = (1 - Z \sin \rho / \rho)^{-1}$$

and

$$\begin{aligned} G &= 1 + Z G_0 \sum_{m=0}^{\infty} (m + \frac{1}{2}) c_{m0} P_m(\cos \psi) \\ &= 1 + \frac{Z e^{i\rho \cos \psi}}{1 - Z \frac{\sin \rho}{\rho}} \end{aligned}$$

The coefficient of Z^n is

$$\Phi(\rho, \psi; n) = e^{i\rho \cos \psi} \left(\frac{\sin \rho}{\rho} \right)^{n-1},$$

which is not independent of ψ but contains the factor $e^{i\rho \cos \psi}$, due to the fact that one step of the chain must always lie along the z axis. The same result can, of course, be got by simpler methods.

The distribution in the general case is again difficult to obtain in closed form for all n from (13.5), though exact expressions for the

moments of r^2 are derivable (see the remark at the end of §8). Approximations can, however, be found as before for long chains. For this purpose it is best to treat

$$L = pG(\rho, \psi; e^{-\nu}) \quad (14.1)$$

as the Laplace transform* of Φ with respect to n , and to seek an asymptotic expansion for large n by the methods of the preceding section, with the following justification. Since

$$\Phi(\rho, \psi; n) = \frac{1}{2\pi i} \int_C G(\rho, \psi; Z) \frac{dZ}{Z^{n+1}},$$

the contour C enclosing the origin but no singularity of G , we also have

$$\Phi(\rho, \psi; n) = \frac{1}{2\pi i} \int_{C'} G(\rho, \psi; e^{-p}) e^{pnd} p,$$

where C' is an open contour in the p plane joining points on $I(p) = \pm \pi$ and having on its left all the singularities of G in the strip. By a suitable choice of C , C' can be made to pass to the right of the origin but to end in the second and third quadrants. On making the change of scale $p' = pN$, where N is a typical (large) value of n , the new contour in the p' plane approximates to the modified Bromwich contour; the effect of replacing it by the latter is to introduce terms which are exponentially small and so do not enter into the asymptotic expansions.

The equations (13.5) now take the form, analogous to (12.3),

$$\left. \begin{aligned} -2pe^p &= L_0(c_{00} - 2e^p) + L_1g_1c_{01} + L_2g_2c_{02} + \dots \\ 0 &= L_0c_{01} + L_1(g_1c_{11} - \frac{2}{3}e^p) + L_2g_2c_{12} + \dots \\ 0 &= L_0c_{02} + L_1g_1c_{12} + L_2(g_2c_{22} - \frac{2}{6}e^p) + \dots \\ &\dots \end{aligned} \right\} \quad (14.2)$$

and the formal solution is

$$L_s = -2pe^p \frac{D_{0s}^\infty}{D_\infty^\infty}, \quad (14.3)$$

where, if $l_{ms} = g_sc_{ms}$ ($s \neq m$), $l_{ss} = g_sc_{ss} - \frac{2}{(2s+1)}e^p$,

$$D_s = \begin{vmatrix} l_{00} & \dots & l_{0\nu} \\ \vdots & & \vdots \\ l_{\nu 0} & \dots & l_{\nu\nu} \end{vmatrix},$$

and $D_{0s,\nu}$ is the cofactor of l_{0s} in D_ν .

* The form (14.1) arises immediately if the Laplace-Stieltjes transform is used.

The existence of the limiting form (14.3), at least for large n , and the order of the approximation when the determinants are replaced by D_{0n} , and D_n , can be established in the following way. It will be sufficient to consider L_0 :

For large n we again seek approximation for small p and ρ with $p = O(\rho^2)$, and if N is a typical value of n , p and ρ^2 may be taken as $O(N^{-1})$. Consider the order of magnitude of l_{ms} . From (13.6) it appears that $c_{ms} = O(N^{-\frac{1}{2}(m+s)})$. If $g(\alpha)$ is not degenerate, $g_0 = 1$, $|g_s| < 1$ ($s \geq 1$); hence $l_{ms} = O(N^{-\frac{1}{2}(m+s)})$, provided m and s are not both zero, but l_{00} is anomalous since

$$l_{00} = c_{00} - 2e^p = -\frac{1}{3}\rho^2 - 2p + \dots = O(N^{-1}).$$

Also l_{ms} is bounded for all m and s , and $l_{ns} = O(1)$ for $s \geq 1$. Hence on dividing the appropriate columns of D_{0n} and D_n by l_{11} , l_{22} , l_{33} , ... convergent determinants are obtained and the existence of the limit (14.3) follows for large N . Moreover, from Schweins's theorem on the ratio of two determinants [Aitken (4)],

$$\frac{D_{00, \nu+1}}{D_{\nu+1}} = \frac{D_{00, \nu}}{D_{\nu}} + g_{\nu+1} \frac{D_{0\nu+1, \nu+1}^2}{D_{\nu} D_{\nu+1}}. \quad (14.4)$$

By extracting factors from rows and columns so as to reduce diagonal elements to terms which are $O(1)$, it can be shown that the second term in (14.4) is $O(N^{-\nu+1})$. It follows that

$$L_0 = -2pe^p \frac{D_{00, \nu}}{D_{\nu}} + O(N^{-\nu}), \quad (14.5)$$

and in a similar way we find

$$L_s = -2pe^p \frac{D_{0s, \nu}}{D_{\nu}} + O(N^{-\nu+\frac{1}{2}(s-1)}), \quad s \geq 1. \quad (14.6)$$

Successive terms of the asymptotic expansion for large N may then be calculated as before.

15. The first approximation is

$$L_0 = \frac{p}{\left(p + \frac{1}{3} \frac{(1+g_1)}{(1-g_1)} \rho^2\right)} + O(N^{-1})$$

$$L_1 = \frac{i\rho}{(1-g_1)} L_0 + O(N^{-1}), \quad L_2 = O(N^{-1}),$$

showing the limiting distribution to be again spherically normal, centred at $(0, 0, a/(1-g_1))$ with the variances of x, y, z equal to $\sigma^2 = \frac{1}{3} \frac{(1+g_1)}{(1-g_1)} na^2$ (cf. Eyring (1932), Moran (1948)).

The calculation of higher order approximations entails much tedious algebra, but the formulae are simplified by the change of scale

$$r^2 = \frac{1}{n} \frac{(1+g_1)}{(1-g_1)} nR^2, \quad P^2 = \frac{1}{n} \frac{(1+g_1)}{(1-g_1)} n\rho^2.$$

The next approximation to Φ_s is then found to be

$$\Phi_0 = \left\{ 1 + \frac{g_1}{(1-g_1^2)} \frac{P^2}{n} + \left[-\frac{1}{2n} \frac{(3+4g_1+3g_1^2)}{(1-g_1^2)} + \frac{1}{10} \frac{(1+g_1)}{(1-g_1)} \right] \frac{P^4}{n} \right\} e^{-\frac{1}{2}P^2}$$

$$\Phi_1 = \sqrt{\frac{3}{n(1-g_1^2)}} \left\{ iP + \left[\frac{2}{5} \frac{(1+3g_1+g_1^2)}{(1-g_1^2)} - \frac{1}{5} \frac{(1+g_1)}{(1-g_1)} \right] \frac{iP^3}{n} \right. \\ \left. + \left[-\frac{1}{2n} \frac{(3+4g_1+3g_1^2)}{(1-g_1^2)} + \frac{1}{10} \frac{(1+g_1)}{(1-g_1)} \right] \frac{iP^5}{n} \right\} e^{-\frac{1}{2}P^2} + O(N^{-\frac{3}{2}})$$

$$\Phi_2 = -\frac{1}{n(1-g_1)} P^2 e^{-\frac{1}{2}P^2} + O(N^{-2})$$

and the corresponding distribution to this order is

$$\frac{1}{\sqrt{2\pi}} R^2 \sin \theta dR d\theta e^{-\frac{1}{2}R^2} \\ \cdot \left\{ \left(1 - \frac{9}{4n} \frac{(1+g_1^2)}{(1-g_1^2)} + \frac{3}{2n} \frac{(1+g_2)}{(1-g_2)} + \left[\frac{(3+2g_1+3g_1^2)}{2(1-g_1^2)} - \frac{(1+g_2)}{(1-g_2)} \right] \frac{R^2}{n} \right. \right. \\ \left. + \left[-\frac{1}{2n} \frac{(3+4g_1+3g_1^2)}{(1-g_1^2)} + \frac{1}{10} \frac{(1+g_2)}{(1-g_2)} \right] \frac{R^4}{n} \right) \\ + P_1(\cos \theta) \sqrt{\frac{3}{(1-g_1^2)}} \left(\frac{R}{n^{\frac{1}{2}}} + \left[-\frac{(13+4g_1+13g_1^2)}{4(1-g_1^2)} + \frac{5}{2} \frac{(1+g_2)}{(1-g_2)} \right] \frac{R}{n^{\frac{3}{2}}} \right. \\ \left. + \left[\frac{(17+16g_1+17g_1^2)}{10(1-g_1^2)} - \frac{9}{5} \frac{(1+g_2)}{(1-g_2)} \right] \frac{R^3}{n^{\frac{5}{2}}} \right. \\ \left. + \left[-\frac{1}{2n} \frac{(3+4g_1+3g_1^2)}{(1-g_1^2)} + \frac{1}{10} \frac{(1+g_2)}{(1-g_2)} \right] \frac{R^5}{n^{\frac{7}{2}}} \right) \\ \left. + P_2(\cos \theta) \cdot \frac{1}{(1-g_2)} \cdot \frac{R^2}{n} + O(N^{-2}) \right\}, \quad (15.1)$$

where R^2 is to be replaced by $\frac{3}{n} \frac{(1-g_1)}{(1+g_1)} r^2$.

It has been checked that for the limiting continuous chain where $t = na$, $E(\alpha^2) \sim 4\lambda a$, $g_1 \sim 1 - 2\lambda a$, $g_2 \sim 1 - 6\lambda a$, (15.1) agrees with (12.6). The terms in P_1 and P_2 do not enter into the radial distribution;

setting $g_1 = g_2 = 0$ for a freely linked chain, we find for the distribution of R ,

$$\sqrt{\frac{2}{\pi}} R^2 dR e^{-\frac{1}{2} R^2} \left\{ 1 - \frac{3}{4n} + \frac{1}{2} \frac{R^2}{n} - \frac{1}{20} \frac{R^4}{n} + O(N^{-2}) \right\}.$$

with $R^2 = \frac{3r^2}{na^2}$, in agreement with Rayleigh's result to this order.

For a chain whose links have complete conical freedom, α has a fixed value and $g_n = P_n(\cos \alpha)$. The particular value $\cos \alpha = \frac{1}{3}$ ($g_1 = \frac{1}{3}$, $g_2 = -\frac{1}{3}$) has a physical application. The radial distribution is

$$\sqrt{\frac{2}{\pi}} R^2 dR e^{-\frac{1}{2} R^2} \left\{ 1 - \frac{33}{16n} + \frac{7}{4} \frac{R^2}{n} - \frac{1}{80} \frac{R^4}{n} + O(N^{-2}) \right\},$$

$$R^2 = \frac{3r^2}{2na^2},$$

which should represent the distribution of end separation of long paraffin molecules in random motion under conditions where conical freedom may be assumed (Treloar, 1949, p. 46). Approximations of this type are, of course, not valid for calculations of entropy at high values of R .

REFERENCES TO LITERATURE

- AITKEN, A. C., 1948. *Determinants and Matrices*, Oliver & Boyd.
- BARTLETT, M. S., 1949. "Some Evolutionary Stochastic Processes", *Journ. Roy. Stat. Soc., B*, xi, 211-229.
- EYRING, H., 1932. "The Resultant Electric Moment of Complex Molecules", *Phys. Rev.*, xxxix, 746-748.
- GOUDSMIT, S., and SAUNDERSON, J. L., 1940. "Multiple Scattering of Electrons", *ibid.*, lvii, 24-29.
- KAC, M., 1949. "On Distributions of Certain Wiener Functionals", *Trans. American Math. Soc.*, lxxv, 1-13.
- McLACHLAN, N. W., 1947. *Theory and Application of Mathieu Functions*, Oxford.
- MORAN, P. A. P., 1948. "The Statistical Distribution of the Length of a Rubber Molecule", *Proc. Camb. Phil. Soc.*, xlv, 342-344.
- MOYAL, J. E., 1950. "The Momentum and Sign of Fast Cosmic Ray Particles", *Phil. Mag.*, xli, 1058-1077.
- PERRIN, F., 1928. "Étude Mathématique du Mouvement Brownien de Rotation", *Ann. Sci. Ec. Norm. Sup.*, xlv, 1-51.
- RAYLEIGH, LORD, 1919. "On the Problem of Random Vibrations and of Random Flights in One, Two and Three Dimensions", *Phil. Mag.*, xxxvii, 321-347.
- ROSSI, B., and GREISER, K., 1941. "Cosmic Ray Theory", *Rev. Mod. Phys.*, xiii, 240-309.
- TRELOAR, L. R. G., 1949. *The Physics of Rubber Elasticity*, Oxford.

XXII.—Artificial Holograms and Astigmatism.* By **G. L. Rogers**, M.A., Ph.D., Department of Physics, University College, Dundee. *Communicated by Professor G. D. PRESTON.*
(With One Plate and Four Text-figures.)

(MS. received August 14, 1951. Revised MS. received November 22, 1951. Addendum received December 14, 1951. Read December 3, 1951)

SYNOPSIS

Experiments in diffraction microscopy, previously described, are here continued. Special emphasis is now laid on verifying the theory by the production of an "artificial" hologram, by non-diffractive means, from data calculated for a relatively simple object. The assumed object is then reconstructed in the usual apparatus.

A type II linear zone plate of limited width is studied as a particular case of an artificial hologram. It gives rise to an unexpected black artefact, which is explained by a detailed analysis of this particular zone plate, and is shown to be due to its limited extent.

Experiments on twisting the linear zone plate skew to the reconstructing beam show that the effective focal length is affected astigmatically by a factor proportional to $\cos^2 \theta$, where θ is the angle of twist, for lines parallel to the axis of twist. Lines perpendicular to the axis of twist are unaffected.

The production of a hologram in an astigmatic pencil and its subsequent reconstruction while skew to a parallel beam is described. It is found that the focal length differences can be corrected in this way, but that the lateral scale factors are only partially rectified.

1. INTRODUCTION

IN an earlier paper (Rogers, 1951) a number of experiments were described in diffraction microscopy in which holograms were prepared by diffraction from a variety of objects, and reconstructed images subsequently obtained under different conditions. It was felt that the theory of the method would be strengthened if it were possible to calculate the shape and intensity of a simple hologram and construct it by quite other means than diffraction, subsequently obtaining an image of the assumed object in the normal way. The first part of the paper deals with attempts of this kind. The second describes some observations on astigmatic effects in diffraction microscopy which followed at once from the first experiments.

2. BLACK-AND-WHITE ARTIFICIAL HOLOGRAMS

In a sense, an ordinary zone plate may be regarded as the calculated hologram of a scattering point, with a rough approximation to the intensity

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

variation in the abrupt black to white transitions of the zone boarders. The results of the work with the zone plate are so well known that it is not proposed to discuss it further here.

The next simplest object is a scattering line, and the linear zone plates considered in I are the appropriate black-and-white artificial holograms of objects of this type (with varying phase-shifts in the scattering line). A fairly detailed theory has already been given, but the experimental work was incomplete at the time of going to press, and is here briefly discussed.

A finite type II zone plate was produced, that is, one based on the series

$$v^2 = 1\frac{1}{2}, 3\frac{1}{2}, 5\frac{1}{2}, \dots 2n - \frac{1}{2} \dots,$$

where v is the distance along Cornu's spiral corresponding to the zonal edges as seen from the focal position. It has a central white zone, with eight black zones and seven white zones on each side, and an extensive exterior region of white, corresponding to a vector from the last v point ($v^2 = 31\frac{1}{2}$) to the convergence point ($\frac{1}{2}, \frac{1}{2}$) of the Cornu spiral. We assume for this discussion that the Cornu spiral has a convergence point, though C. L. Andrews (1951) has challenged this assumption. It is possible that a careful study of the theory of finite linear zone plates might decide the issue.

The linear zone plate behaved entirely as predicted in the first order, giving the usual wave-length variation. The theoretical focal length is obtained from the dimensions of the zone plate quite easily. The distance, x , from the centre line to any zone boundary is related to the v of the Cornu spiral by the relationship:

$$x = v\sqrt{\left(\frac{1}{2}f\lambda\right)},$$

where f is the focal length of the zone plate to wave-length λ .

Transforming this, we get

$$f = \frac{2}{\lambda} \left(\frac{x}{v} \right)^2.$$

A set of values for x for the edges of the zones was obtained by travelling microscope, and these were squared and divided by the appropriate values of v^2 , viz. $1\frac{1}{2}$, $3\frac{1}{2}$, $5\frac{1}{2}$, etc. The value of $(x/v)^2$ thus obtained was reasonably constant, there being a slight indication that the dark zones were a little larger than they should be, possibly due to lateral spread of the light during printing. There was also a suggestion that the central zone was a few per cent. too narrow, of which more later. The average value was $(3.11 \pm 0.05) \times 10^{-8}$ sq. cm., giving f for the mercury green equal to 114 cm.

The measured value was 115 cm., in better agreement than we have a right to expect.

Because of the finite size of the zone plate high-order effects are not to be expected, as we shall see. But an important anomaly was observed. A *black* line was formed in the centre of the field when the eyepiece was near to the third-order position. It was, however, clear that the position of this anomaly was not exactly *at* the third-order position, but was a little outside it. That is, it corresponds to an order slightly less than three.

3. THEORY OF THE ANOMALOUS DARK LINE

It was therefore decided to explore this phenomenon in theory. In order to do so, we develop further the theory of the Cornu spiral given in I. The analysis is ultimately graphical, the central vector being plotted on the Cornu spiral, and those due to the outer zones being calculated in magnitude and direction, and being drawn in to form the vector polygon. A sample polygon is given in fig. 1.

It is shown in I that the slope and radius of curvature of the Cornu spiral are readily obtained from its intrinsic equations $s = v$, $\psi = \frac{1}{2}\pi v^2$, giving $\rho = 1/\pi v$. We can therefore transform to polar co-ordinates centred on $(\frac{1}{2}, \frac{1}{2})$ for all portions of the spiral over $v=2$, as the point $(\frac{1}{2}, \frac{1}{2})$ is then always close to the centre of curvature of the spiral. The lengths of vectors in the outer zones can readily be calculated in terms of these polar co-ordinates to an accuracy of a few per cent. The initial line of the polar co-ordinates is conveniently taken parallel to the x or $C(v)$ direction.

Now a high-order effect is obtained by approaching the zone plate more closely than its normal focal distance, whereby the fixed markings subtend a larger angle at the new point of observation, and thus correspond to positions further from the centre of the Cornu spiral, as drawn from this new point of observation. If the new point of observation is at a distance f/n from the zone plate, f being its primary focal length, the zone boundaries will now occur at values of v on the new Cornu spiral given by

$$v^2 = n \times (1\frac{1}{2}, 3\frac{1}{2}, 5\frac{1}{2}, \dots 2n - \frac{1}{2}, \dots).$$

If n is an odd integer we may expect unusually great effects in some cases, corresponding to high-order foci on an ordinary zone plate. But for the purpose of this analysis we shall assume that n can have any value.

In our particular case, the contribution from the inner zone is obtained graphically, and that from $v^2 = n \times 1\frac{1}{2}$ to $v^2 = n \times 3\frac{1}{2}$ is blacked out. The first outer zone is thus that from $v^2 = n \times 3\frac{1}{2}$ to $v^2 = n \times 5\frac{1}{2}$. The vector to

which this can give rise can be obtained at once by solving the appropriate triangle in polar co-ordinates. This procedure is repeated for each outer vector, until the vector from $v^2 = \pi \times 31\frac{1}{2}$ to the point $(\frac{1}{2}, \frac{1}{2})$ is reached. This vector is obtained from the polar equation directly.

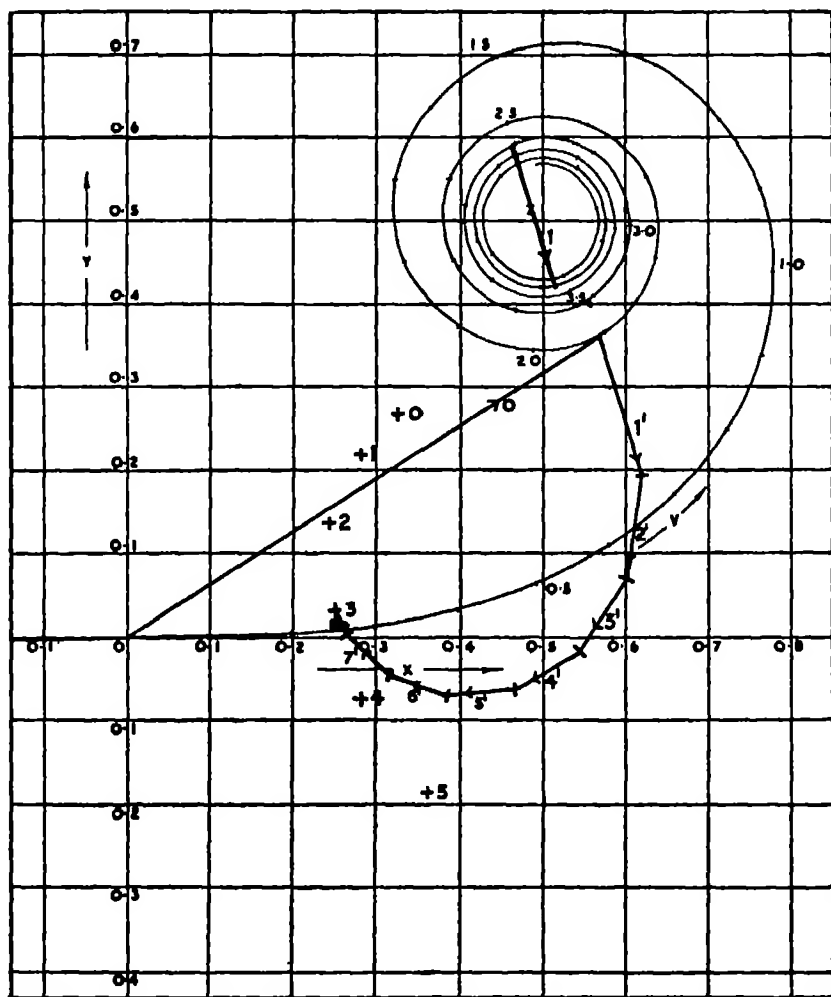


FIG. 1.—Cornu's spiral.

In order to search for an anomaly in the region just below $n=3$, calculations have been made for values of n of 2.90, 2.91, 2.92, 2.93, 2.94 and 2.95. It was found that the lengths of the outer vectors, including the final vector to $(\frac{1}{2}, \frac{1}{2})$, did not change at all rapidly in this region, and could be represented to better than $\frac{1}{2}$ per cent. by the values for $n=2.925$. The main change occurs in the angular arrangement of these vectors, which is

regular one to another. The external angles of the several polygons vary uniformly from 36° for $n=2.90$ to 18° for $n=2.95$ (and is, of course, zero for $n=3.0$). The final vector, E , representing the open field from $v^2=n \times 31\frac{1}{2}$ to infinity, can be shown to subtend an external angle at the end of the polygon which is exactly $\frac{1}{2}$ of that ruling up to that point.

The polygons corresponding to these six cases were plotted from the data thus obtained, and the ends of the polygons carefully noted. Fig. 1 shows one such polygon for $n=2.93$. The central zone vector is marked o , and is drawn from the origin to the point where $v^2=n \times 1\frac{1}{2}$. The outer vectors have been drawn in and suitably numbered. The final point, at the end of the vector E , is numbered 3. The end points of the other five polygons have been transferred to this diagram and marked o , 1, 2, 4 and 5 respectively. It will be seen that the end points sweep out a smooth curve with increasing n , and that this curve ϕ passes close to the origin at $n=2.93$.

The darkness thus produced does not look very marked on the Cornu spiral, but this is an amplitude diagram, and the resultants must be squared to give the intensity. Moreover, this calculation is for the ideal zone plate. If the actual zone plate has too narrow a central open space, this will contract the vector o a little, and thus move the polygon bodily to the left, still further reducing the gap with the origin. It is thought that the depth of the black line produced by this zone plate owes something to the slightly contracted core.

If the total length of the vectors is summed without regard to phase, it is found to be $0.75 v$ units. This is by no means sufficiently large to swamp the initial o vector, as is required by the theory of paper I, even if the vectors are dead in line, as they will be at order 3.0 . We do not, therefore, expect to get any very marked enhancement of the intensity at $n=3.0$, and for higher orders the vector sum of the outer zones of this particular zone plate will be still less, and will therefore produce less and less effect on the initial vector o . This explains the observed absence of high-order effects.

4. FINITE LINEAR ZONE PLATES

It is of interest to consider here the effects likely to be produced by a finite linear zone plate, with a view to estimating the number of orders likely to be obtained from a given number of zones. It is convenient to take a type I zone plate, with zone edges given by the v values where

$$v^2 = 1, 3, 5, 7, \dots, \text{etc.}$$

It was shown in I that in this case the sum of the outer zones (leaving aside the central one) is

$$S_{\text{outer}} = \frac{1}{\pi} \left\{ \frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{4}} + \dots, \text{etc.} \right\}.$$

A lower limit to this sum is

$$\frac{1}{\pi} \left\{ 1 + \frac{1}{\sqrt{2}} \cdot \frac{(\sqrt{2})^k - 1}{\sqrt{2} - 1} \right\},$$

obtained by grouping 2^k terms of the first series into k sub-groups of geometrically decreasing value (approximately).

This sum refers to the first order, but it is shown in I that its value in the n th order will be $1/\sqrt{n}$ of the above. This n th order vector is to be comparable with the central zone vector, whose length is roughly $1/\sqrt{2}$.

On this basis, we find that a zone plate with 64 outer zones on each side would give orders up to $n=34$, while one of 8 outer zones each side goes up to $n=3.4$. The latter result is in reasonable agreement with the last section.

5. MORE COMPLEX OBJECTS

So far the objects treated have been very simple. The next object tried was a finite wire or slit obstruction, calculations for which were available in routine optical lecture notes. The calculations were not very suitable: the hologram still closely resembled the original object, and the number of external "fringes" available was only three. The reconstructions obtained are not regarded as very satisfactory.

The problem of producing a continuous tone hologram was next studied. A set of calculations from the Fresnel Integrals was kindly done by Mr J. W. B. Laing for an object consisting of two wires or slats, of equal breadth, separated by a distance slightly (20 per cent.) wider than their breadths. The calculation was in a plane such that the double structure of the object was entirely obscured by diffraction, the pattern looking like that of a single obstruction to the untutored eye.

The intensities thus obtained were plotted on a graph of I against the parameter v from the spiral. The plot was repeated below the v -axis to form a double symmetrical plot, and the outer regions were blacked out to form a pattern similar to a Phillips-Miller type variable area sound track (Wood, 1940). The conversion of this to a variable density record is readily achieved by a well-known dodge. The plot A (fig. 2) is set up with its I -axis vertical, and is uniformly illuminated. An image is formed on

the plate BD by a cylindrical lens, C , with its cylindrical axis, LL' , vertical, and its optic axis OO' (if one may call it such) horizontal and perpendicular to BD and A . The line BD is horizontal through O' , and the centre of symmetry of A is at O . The intensity of light falling on any point P of BD is proportional to the height of the white element of area QR , which

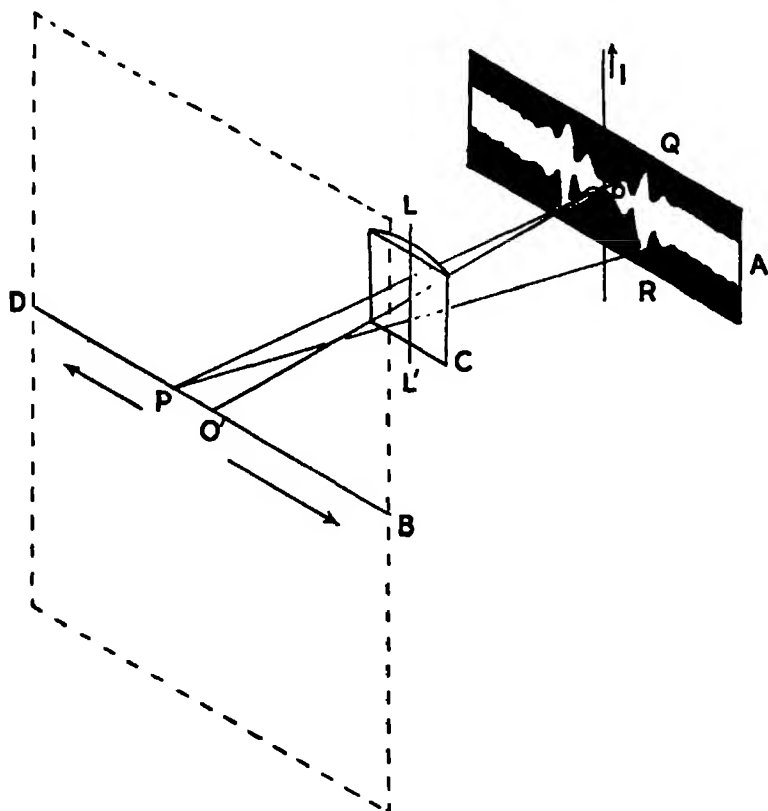


FIG. 2

corresponds to it optically in the cylindrical axis LL' of the cylindrical lens C (which axis is equivalent to the centre of a thin spherical lens: rays go through it without deviation). This assumes that the lens is sufficiently high to take in the extreme rays PQ , PR . In practice, it was arranged that the lens was ample, with a fair margin, so that a moderate strip of about $\frac{1}{4}$ in., centred on BD , was all essentially of the right density for the variable density record. When printed with a contrast of 2, this is of course equivalent to a hologram.

A cylindrical lens was taken from an optician's test set and stopped down laterally with a slit $\frac{1}{4}$ in. wide (retaining its full height of 1 in.). It was

then mounted on the copying camera, in place of the usual spherical lens, with the cylinder axis vertical, and the optic axis horizontal. The plot was set up as above, and the focusing was done in blue light with a filter. The exposure was then made on an ordinary plate, sensitive to the blue only. This reduces chromatic aberration, and fits in with our standard processing technique. A step wedge is subsequently imposed on one edge, to control the contrast.

In order to ensure that only the central portion was used, prints were made on paper, and the unwanted portions removed with a guillotine. Four such strips were pasted together in order to give the necessary height. It was then further reduced in size photographically to provide the hologram finally used.

Fiducial marks were carried through the operation to indicate the length equivalent to 22.0ν units, and a measurement of these on the final hologram gives us the focal length, as with the linear zone plate. The theoretical focal length in mercury green is 72.7 cm. The hologram reconstructs the theoretical double slit pattern very well, at a focal length to mercury green of 69 ± 6 cm. The agreement is quite reasonable. The Plate gives an outline of the steps in the process, and the final result.

6. ASTIGMATISM

During this work we became aware of the bearing of this type of linear pattern on the problem of astigmatism. If a linear zone plate be arranged with its length vertical and its plane perpendicular to a horizontal beam of parallel light, it produces the normal bright line image in its focal plane. If, now, the hologram or zone plate be twisted about a vertical axis through an angle θ , so that its plane is no longer perpendicular to the horizontal beam, it acts as a cylindrical lens of higher power. It will produce a line focus at a distance $f \cos^3 \theta$, where f is the normal focal length. If, however, the zone plate is originally horizontal, and not vertical, its focal length is unaffected by twisting about a vertical axis. Hence we may expect that with a generalized pattern, twisting the hologram will introduce an element of astigmatism of amount proportional to $\cos^3 \theta$. Contrariwise, twisting a hologram with astigmatism present, such as may easily arise in the electron case, might conceivably be used to remove it.

The $\cos^3 \theta$ law is to be expected on theoretical grounds. The twisted zone plate may be regarded as equivalent to a finer zone plate of shorter focus, obtained by projecting it on to the plane normal to the light beam.

The change of linear scale produced by this projection is $L = \cos \theta$. It is shown in I that $f \propto L^3$, which explains the observations.

A careful set of measurements is shown in graphical form in fig. 3.

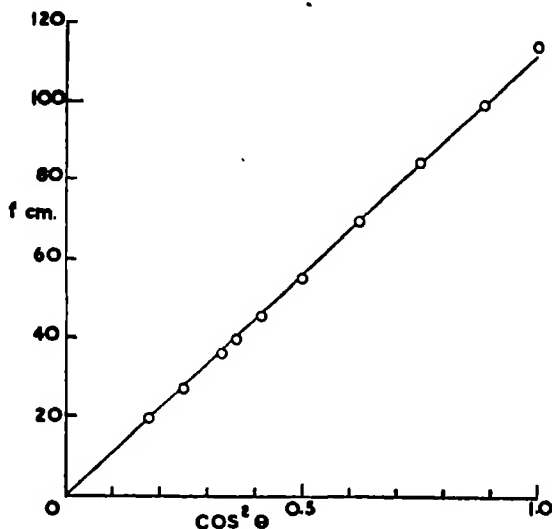


FIG. 3

7. PRODUCTION OF A HOLOGRAM IN AN ASTIGMATIC BEAM

In order to test the possibility of correcting an astigmatic hologram by mounting it skew to the incident light, it was necessary to produce an astigmatic hologram. A cylindrical lens was therefore placed in the filter-holder of the apparatus described in I, and this, together with the usual microscope objective, in this case of 1 in. focus, produced a much reduced image of the fine hole in front of the arc. The image was no longer a point image, but consisted of two focal lines, vertical and horizontal, very close together. The cylindrical lens was unfortunately not very powerful compared with the microscope objective, and hence the separation was not very good.

In order to get appreciable astigmatism in the hologram, the object had to be placed very close to the astigmatic pencil. The position of the vertical line, the horizontal line and the object were determined by a "location run" as described in I, but unfortunately this is not a very good method, and the main uncertainties of the work lie here.

As object, a very much reduced picture of three fan-like devices was used, two fans lying vertically and one horizontally. They were designed to aid

precise focusing in two directions at right angles. The distance between the two vertical fans, a , was 1.455 times the distance, b , of the horizontal fan (tip) from this line. It is, of course, important to keep track of this ratio throughout the experiment.

A hologram was taken in mercury blue light (4358 Å.) with the vertical line 8.75 cm. from the plate, the horizontal line 8.62 cm. and the object 8.15 cm. from the plate. The powers in dioptries in blue light calculated from these figures are 0.90 dioptries for the vertical line and 0.66 for the horizontal line. Unfortunately, these powers are only obtained by differencing comparatively large numbers, of the order of 11 m^{-1} , and hence must be regarded as indicative only. Converted to mercury green they become 1.13 and 0.83 dioptries respectively.

Straight reconstructions were made in green light using the hologram normal to the beam, and the observed powers were found to be 1.27 and 0.96 dioptries respectively, in better agreement than we might expect. The hologram was then gradually twisted, and the best position judged by eye. It was decided that a good focus was obtained for all three fans when the deflection was 29° . The two fans were horizontal, as the power for the two fans (originally vertical in taking) is higher than for the one fan, and it is the latter which must be strengthened by twisting. The angle 29° implies a power ratio of $1:0.76496$ or, say, $1:0.765$. The ratio from the taking data is $1:0.733$ and from the reconstruction focal lengths $1:0.755$. The agreement must be regarded as satisfactory.

8. LATERAL SCALE DISTORTION

In any work with astigmatic beams, the possibility that the scale or magnification of the image might be different in the two principal directions is a very real one. For instance, in taking the hologram, the projective magnification in a horizontal plane, due to the vertical line, is $\times 13.7$, whereas that vertically due to the horizontal line (which is nearer the object) is $\times 18.3$, giving an exaggeration in a vertical direction. Thus the original vertical/horizontal ratio of 1.455 is expected to come out at $1.455 \times \frac{18.3}{13.7} \approx 1.95$. From the hologram we get the ratio about $2:1$, this being difficult to judge as the detail of the fan points is lost by diffraction. The straight reconstructions both confirm this ratio of $2:1$.

The most interesting ratio is that of the compensated reconstruction. This gives a ratio of $1.77:1$, which is very nearly $2 \cos 29^\circ:1$, or $1.75:1$. It is clear, therefore, that the skew projection produces a correction to the

scale factor of the right sign, but not of the right magnitude. It cannot, therefore, be used for compensating for a hologram involving any serious astigmatism.

On the other hand, the astigmatic effects with this hologram are very marked, and hence it is possible that electron holograms, where the scale distortion and astigmatism is in any case small, could be corrected by this device in a manner acceptable in practice.

REFERENCES TO LITERATURE

- ANDREWS, C. L., 1951. "A Correction to the Treatment of Fresnel Diffraction", *Amer. Journ. Phys.*, XIX, 280-284.
ROGERS, G. L., 1951. "Experiments in Diffraction Microscopy", *Proc. Roy. Soc. Edin.*, A, LXIII, 193-221.
WOOD, ALEXANDER, 1940. *Acoustics*, Blackie & Sons, p. 508, fig. 18.5.

ADDENDUM

From the Department of Physics, Sir John Cass College,
Jewry Street, Aldgate, London, E.C.3

(Received December 14, 1951)

Since taking up a temporary appointment at the Sir John Cass College, the author has had the opportunity of checking the artificial hologram discussed in the main paper on a non-recording microphotometer. The result is interesting in demonstrating both the points of agreement between the actual and theoretical holograms, and the points where the technique of production has failed.

The results are given in fig. 4. The upper curve gives the uncorrected galvanometer deflection as the artificial hologram was scanned: the divisions on the horizontal axis are millimetres. The instrument is said by the manufacturers to give a linear response, but this is difficult to check accurately. Within the limits imposed by the small vertical scale of this diagram, however, the departures are not likely to be serious.

The second curve is the theoretical curve, as calculated, and gives I^2 against units in v .

Comparing these two curves, we note at once that the dark regions of

the artificial hologram are not by any means as dense or obstructive to the incident light as the theoretical curve requires. The contrast of the first light fringes from the centre is exaggerated, that of the second (the real maxima of the pattern) is depressed, and the rest is comparatively well copied. But when the comparatively large distance of the variations from the zero axis is taken into account, it will be seen that the contrast is low compared with the theoretical.

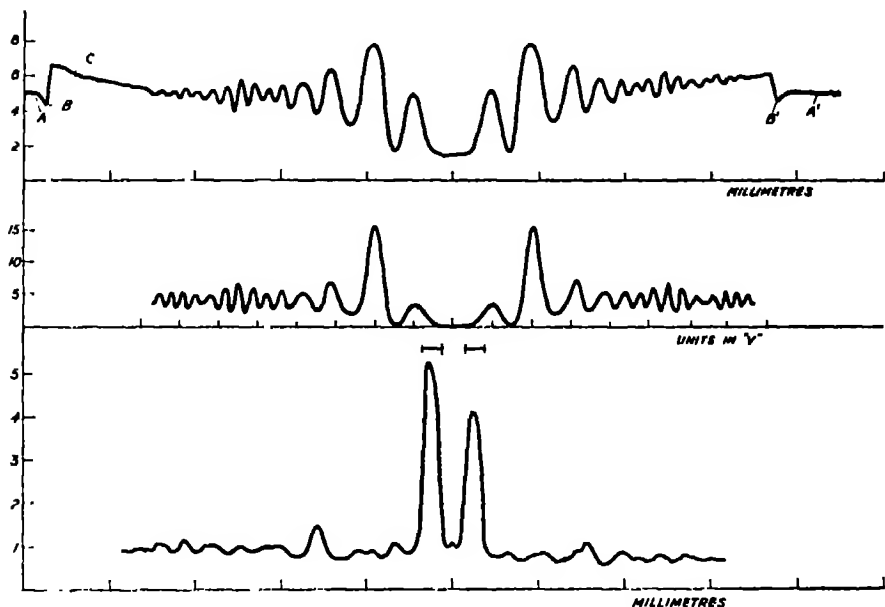


FIG. 4

It is suggested by Gabor (Gabor, 1951) that the high level of this artificial hologram is not serious, but is equivalent to adding a uniform background to the reconstructed wave.

It will also be noticed that at the ends of the regular curves of the top figure, certain irregularities are observed. *A*, *A'* give the level of the card used as the background to the mounted photographic prints. *B*, *B'* are small shadows cast by the edges of the photographic prints, as they lie above the card. The single peak, *C*, is due to slight glare from the surface of the print at one side in this region.

The lowest curve is a microphotometer trace across the *negative* of the reconstructed image. On the negative, the two peaks in transmission correspond to the two opaque bars postulated in the calculations. The two small marks at the top of this diagram give the theoretical position of these bars.

The asymmetry of this plot is not fully understood, but is connected with the fact that the intensity across the reconstructing beam was not uniform. The magnitude of the effect is, however, surprisingly large.

Part of the diffraction pattern of the unwanted secondary object is discernible in the background. The units along the horizontal axis are once again millimetres.

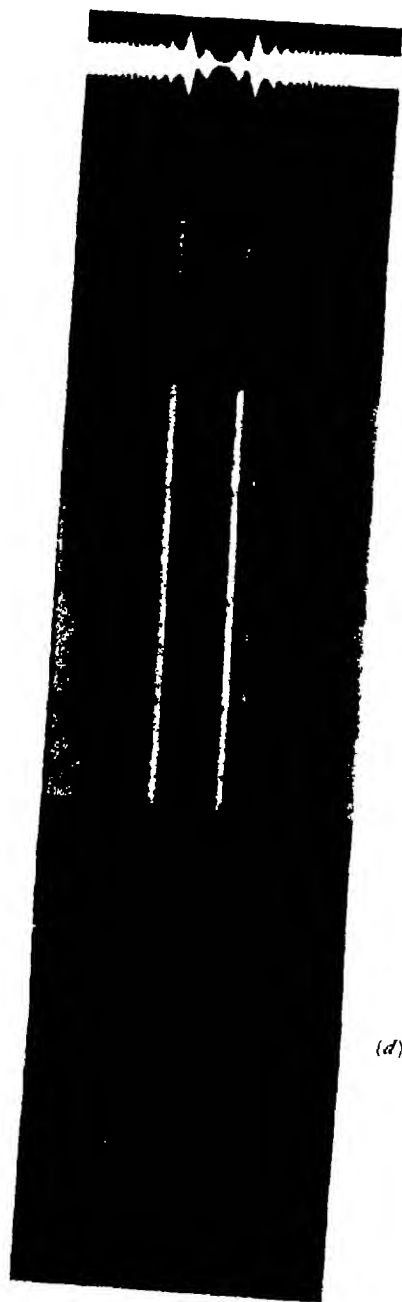
ACKNOWLEDGMENTS

I have to acknowledge the encouragement and assistance of Professor Preston at Dundee, and of Mr R. H. Humphry at the Cass College. I am also grateful to the Principal of the Cass College for arranging a temporary appointment and giving me laboratory space during a gap in my normal employment.

REFERENCE

GABOR, 1951. Private communication.

(Issued separately August 23, 1952)



(a) Phillips-Miller type plot of calculated intensities

(b) Photo of (a) through a cylindrical lens, with step wedge

(c) Hologram prepared from centre strips of (b) at higher contrast

(d) Reconstruction from (c) of the original theoretical pattern.

G. L. ROGERS

[PLATE

XXIII.—Studies in Practical Mathematics. VII. On the Theory of Methods of Factorizing Polynomials by Iterated Division.* By **A. C. Aitken**, D.Sc., F.R.S., Mathematical Institute, University of Edinburgh.

(MS. received November 30, 1951. Revised MS. received January 24, 1952.
Read May 5, 1952)

SYNOPSIS

The division of one polynomial by another is studied with the object of ascertaining the errors produced in the coefficients of successive remainders by small errors in the coefficients of the divisor. It is shown that the matrix which effects this transformation of errors is a polynomial in the rational canonical matrix for which the divisor polynomial is characteristic. The theory gives rise to a numerous class of iterative processes for finding an exact factor, such as the extant method based on the penultimate remainder, *Baird's* iterative method of finding a quadratic factor, and many others. Some new suggestions are made for accelerating convergence.

I. INTRODUCTORY

IN an earlier paper (Aitken, 1950) the author examined the theory of a method first proposed and used by S. N. Lin (Lin, 1939, 1941) for approximating by repeated division to an exact factor of a given polynomial. This was called the method of the *penultimate remainder* (p.r.). It is shown below that there are many such methods, though that of the p.r. would seem to be the simplest; and occasion is taken to give a comprehensive theory for these methods.

Note.—It calls for remark that Dr Lin almost immediately superseded the method of the *penultimate remainder* (not so named by him) in favour of a second and in general more rapidly convergent method (Lin, 1941, p. 241, and 1943) which stands to the former very much as Seidelian iteration for the solution of simultaneous linear equations stands to simple iteration. It is this later method that now bears the name of Lin's method. The theory of its convergence is less straightforward than that of *penultimate remaindering* and will be reserved for a separate communication.

2. THE TRANSFORMATION OF SMALL ERRORS IN REMAINDERS

It is well known and evident, yet important for the present purpose, that *remaindering* with respect to an arbitrary divisor, say $d_m(x)$, is a *linear* operation, in the sense that if $f(x)$ and $F(x)$ are arbitrary polynomials

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

which are being divided by $d_m(x)$, and if respective remainders $r_h(x)$ and $R_h(x)$ of the same degree h are reached at a certain stage (not necessarily the final stage) in these divisions, then the remainder of degree h arising from division of $\lambda f(x) + \mu F(x)$ by $d_m(x)$ is $\lambda r_h(x) + \mu R_h(x)$.

Let, then, the polynomial

$$f_n(x) \equiv x^n - a_1 x^{n-1} + a_2 x^{n-2} - \dots + (-)^n a_n \quad (1)$$

$$= (x - a_1)(x - a_2) \dots (x - a_n) \quad (2)$$

have an exact divisor $d_m(x)$ of degree $m < n$, where

$$d_m(x) = x^m - b_1 x^{m-1} + b_2 x^{m-2} - \dots + (-)^m b_m \quad (3)$$

$$= (x - \beta_1)(x - \beta_2) \dots (x - \beta_m), \quad (4)$$

the quotient in this division being

$$q_{n-m}(x) = x^{n-m} - c_1 x^{n-m-1} + c_2 x^{n-m-2} - \dots + (-)^{n-m} c_{n-m}. \quad (5)$$

Suppose, however, that we divide $f_n(x)$ not by $d_m(x)$ but by $t_m(x)$, which, except for small errors ϵ_i in the coefficients of its terms later than x^m , would be the same as $d_m(x)$; then errors will be produced in the coefficients of the successive remainders. To find these, let $x^s d_m(x)$, where s is an arbitrary integer, be divided by

$$t_m(x) = x^m - (b_1 + \epsilon_1)x^{m-1} + (b_2 + \epsilon_2)x^{m-2} - \dots + (-)^m(b_m + \epsilon_m). \quad (6)$$

The first remainder is evidently

$$\epsilon_1 x^{s+m-1} - \epsilon_2 x^{s+m-2} + \dots + (-)^{m-1} \epsilon_m x^s. \quad (7)$$

We shall associate its coefficients with the vector of errors in the coefficients b_i of $t_m(x)$, namely $\epsilon = \{\epsilon_1 \epsilon_2 \dots \epsilon_m\}$. The next remainder in this division, if terms of higher than the first degree in the errors are neglected, is seen to be

$$(b_1 \epsilon_1 - \epsilon_2) x^{s+m-2} - (b_2 \epsilon_1 - \epsilon_3) x^{s+m-3} + \dots + (-)^{m-1} b_m \epsilon_1 x^{s-1}. \quad (8)$$

If, therefore, the error-vector at this second stage is $\{\epsilon'_1 \epsilon'_2 \dots \epsilon'_m\}$, we have

$$B\epsilon \equiv \begin{bmatrix} b_1 & -I & . & . \\ b_2 & . & -I & . \\ b_3 & . & . & -I \\ \vdots & \vdots & \vdots & \vdots \\ b_m & . & . & . \end{bmatrix} \begin{bmatrix} . \\ . \\ . \\ -I \\ . \end{bmatrix} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_m \end{bmatrix} = \begin{bmatrix} \epsilon'_1 \\ \epsilon'_2 \\ \epsilon'_3 \\ \vdots \\ \epsilon'_m \end{bmatrix}. \quad (9)$$

Thus B is the matrix that transforms the error-vector in a remainder at one stage into the error-vector in the next remainder. The error-vector at the k th stage of this particular division will be $B^{k-1}\epsilon$. In the theory of matrices B is familiar as a *rational canonical form*, except that more commonly B' , its transpose, is taken as canonical. Its characteristic polynomial is $(-)^m d_m(x)$, and so its latent roots are the zeros β_i of $d_m(x)$. Its latent vectors are readily seen (in the case of distinct latent roots) to be

$$\left\{ 1 \sum_{[A]} \beta_i, \sum_{[A]} \beta_i \beta_j, \dots, \beta_i \beta_j \beta_k \dots \beta_m \right\}, \quad h=1, 2, \dots, m, \quad (10)$$

where $[h]$ denotes that β_h is supposed in all the indicated summations and in the last product in the bracket. If the elements in these latent vectors are given alternate plus and minus sign, they appear as coefficients in the divisor polynomials $d_m(x)/(x - \beta_h)$ of degree $m-1$, such as figure, for example, in Lagrange's interpolation formula.

3. THE REDUCING TRANSFORMATION

When a remainder of degree m is used iteratively as a divisor, it is convenient to ensure, by dividing through by the coefficient of its highest term, that this term shall become x^m . We therefore examine this *reducing* transformation of errors, produced when a remainder, which for convenience we shall write in the form

$$k(1 + \epsilon_0'')\{x^m - (b_1 + \epsilon_1'')x^{m-1} + \dots + (-)^m(b_m + \epsilon_m'')\}, \quad (1)$$

is divided by its leading coefficient $k(1 + \epsilon_0'')$. Carrying out this division and neglecting, as before, error terms of higher than the first degree, we find that the error-vector has been transformed by

$$-k^{-1} \begin{bmatrix} b_1 & -1 & . & . & . & . \\ b_2 & . & -1 & . & . & . \\ b_3 & . & . & -1 & . & . \\ \vdots & & & & \ddots & \\ b_m & . & . & . & . & -1 \end{bmatrix}. \quad (2)$$

Now this, except for the scalar factor k^{-1} , is merely B , extended consistently by a further column. However, if the remainder is *penultimate* the coefficient in its last term has no error; for it is simply the constant term of the dividend $f_n(x)$ "brought down". Therefore in reducing a *penultimate* remainder the last column in (2) is unnecessary, and the reducing

transformation is $-k^{-1}B$, acting on a vector of m elements. We may note similarly that in an *antepenultimate* remainder the last *two* coefficients would be free from error, and then the reducing transformation would be $-k^{-1}B$ with last column removed. Such reducing operations play an important part in the theory of iterated remainders; and so it will be convenient to use the notations $B(m, m+1)$ and $B(m, m-1)$ for the matrices obtained by extending or diminishing B by a column, and to use a similar notation for the extensions or diminutions, consistently applied, of powers of B . At the same time let us denote the " B " corresponding to a divisor $(x-\nu)d_m(x)$ by B_ν , so that

$$B_{\nu, m} = \begin{bmatrix} b_1 + \nu & -1 & . & . & . & . \\ b_2 + \nu b_1 & . & -1 & . & . & . \\ b_3 + \nu b_2 & . & . & -1 & . & . \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \nu b_m & . & . & . & . & -1 \end{bmatrix}. \quad (3)$$

Then it is easy to verify the following results:

$$\begin{aligned} B(m, m+1)B &= B^2(m, m+1), & B(m, m+1)B^2 &= B^{2+1}(m, m+1), \\ B(m, m+1)B^2(m+1, m) &= B^{2+1}. \end{aligned} \quad (4)$$

The important feature is that $B(m, m+1)$, as a premultiplier, obliterates ν .

4. APPLICATION TO PENULTIMATE REMAINDERING

The above results give the theory of iterated division and remaindering. In particular, for the method of the penultimate remainder, since

$$f_n(x) = \{x^{n-m} - c_1x^{n-m-1} + \dots + (-)^{n-m}c_{n-m}\}d_m(x), \quad (1)$$

the exact penultimate remainder with respect to the divisor $d_m(x)$ is $(-)^{n-m}c_{n-m}d_m(x)$. Now by § 2 (9) and the linear nature of remaindering, the errors in the penultimate remainder with respect to $t_m(m)$ will constitute the error-vector

$$\{B^{n-m-1} - c_1B^{n-m-2} + \dots + (-)^{n-m-1}c_{n-m-1}I\}e, \quad (2)$$

and on this the reducing transformation, by § 3 (2) *et seq.*, will impose the further transformation $(-)^{n-m-1}c_{n-m}^{-1}B$. Thus the vector of errors in the *reduced* penultimate remainder (r.p.r.) will be Re , where

$$R = (-)^{n-m-1}c_{n-m}^{-1}\{B^{n-m} - c_1B^{n-m-1} + \dots + (-)^{n-m-1}c_{n-m-1}B\}. \quad (3)$$

The condition for convergence of the iteration is therefore that all the latent roots of R , namely each

$$\rho \equiv (-)^{n-m-1} c_{n-m}^{-1} \{\beta^{n-m} - c_1 \beta^{n-m-1} + \dots + (-)^{n-m-1} c_{n-m-1} \beta\}, \quad (4)$$

where β is a latent root of B , that is to say a zero of $d_m(x)$, shall be such that $|\rho| < 1$. The rapidity and nature of the convergence will depend on the root or roots of maximum modulus.

These results were implicit in our former paper (Aitken, 1950), but were not clearly perceived as admitting the present formulation. They give a more significant expression to the essential basis of the method of iterated r.p.r., namely, that apart from sign the error-transforming matrix R is a polynomial in B which exactly copies the reduced penultimate quotient.

The conditions of convergence involve the divisor $d_m(x)$, which is not known exactly, and which it is the very object of the method to ascertain. Consequently, the nature and rapidity of the convergence can be determined in practice only approximately; but this is a limitation inherent in every method that proceeds towards an unknown solution by successive approximation.

5. ALTERNATIVE ITERATIVE PROCESSES

At this stage many other less simple iterations, also based on remainders, suggest themselves. For example, we might divide $f_n(x)$ by $t_m(x)$, stopping at the *antepenultimate* remainder, of degree $m+1$; next, taking this, reduced, as a new divisor, we might divide $f_n(x)$ again up to the *ultimate* remainder, which will be of degree m , and reduce. This two-stage process could then be iterated. We find without trouble, by the linearity of remaindering and § 3 (4), that the error-transforming matrix for these two stages together is the product of two polynomials in B , the first copying the reduced antepenultimate quotient in the first division, the second copying the reduced ultimate quotient in the second division. In fact a general theorem, probably only of theoretical interest, may be enunciated at this point.

Let $f_n(x)$ be divided by $t_m(x)$ up to any remainder of degree $< m$. Let the remainder, reduced, be divided again into $f_n(x)$ up to any remainder of degree $< m$. Continue thus, but so that the last division in such a sequence of operations gives a remainder of degree m . Then, under conditions of convergence, such a cycle can be iterated to give, in the limit, $d_m(x)$. The error-vector in the last reduced remainder is in fact Ra , where, with suitable sign, R is a product of polynomials in B which

exactly copy the successive reduced quotients, partial or ultimate as the case may be, that would arise if the divisions were exact.

It is hardly necessary to give a numerical illustration of the process of antepenultimate remaindering. Even though its two-stage cycle may often show more convergence than two consecutive stages of penultimate remaindering, the latter is the simpler process; and modern desk machines, to say nothing of automatic machines working to a coded programme, have so reduced the time and tedium of computation that a simple method such as penultimate remaindering, or Lin's later method, provided that it is reasonably convergent, is to be preferred to a more complicated method, even if the latter is more rapidly convergent. Some have questioned whether penultimate remaindering and Lin's later method are as useful in practice as the time-honoured Newton-Raphson iteration or its extension (Bairstow, 1920) to the case of a quadratic factor; for these can be carried out in the same way, admittedly in *two* stages for each iteration, by Horner processes of division and remaindering. The answer probably is: that for purposes of programming, the method of r.p.r. is the simplest of all, and Lin's method almost as simple. Therefore it will be of value to see whether the convergence of these methods can be accelerated in any convenient way.

6. ACCELERATIVE PROCESSES: QUADRATIC CONVERGENCE

In an earlier paper on this subject (Aitken, 1950) certain devices, already known to be serviceable in other applications, were applied to enhance the convergence of penultimate remaindering. We now survey some other possibilities of the same kind.

Consider first the *ultimate* remainder in the division of $f_n(x)$ by $t_m(x)$. Here we easily find, in the notation and by the theorems of §§ 2 and 3, that the error-vector of the coefficients in the remainder is $q_{n-m}(B)\mathbf{a}$. Thus, in theory, if we operate on these remainder-coefficients with $\{q_{n-m}(B)\}^{-1}$, we shall obtain, apart from terms of higher than the first order, the corrections to be applied to the coefficients of $t_m(x)$. In practice we shall not know $q_{n-m}(B)$ exactly, and therefore shall have to use the approximation to it provided by the work-sheet. The application of this correction, for small values of m , is equivalent to extant methods. For example, when $m=1$, the correction is precisely the classical Newton-Raphson one; when $m=2$, the two correction terms obtained by applying the approximate $\{q_{n-m}(B)\}^{-1}$ to the two coefficients in the remainder are precisely Bairstow's corrections (Bairstow, 1920, pp. 558-560) for the case of a quadratic divisor. For completeness and for later comparison, though there is no

essential novelty here, we shall illustrate by examples the appearance of the work-sheet when two successive divisions are used to effect these corrections.

Example 1.—To improve the approximation $x = 1.75$ for a zero of the fourth Laguerre polynomial $x^4 - 16x^3 + 72x^2 - 96x + 24$.

		1	16.0	72.0	96.0	24.0
1	1.75	1	14.25	47.0625	13.6406	0.1290
	-0.00424	1	12.50	25.1875	-30.4375	
1	1.74576	1	14.25424	47.11552	13.74761	-0.00003

Here the correction, namely $-0.1290/30.4375 = -0.00424$, is the Newton-Raphson one, the divisions being the usual Horner processes for obtaining by the remainder theorem a functional value and a derivative.

Example 2.—With the same Laguerre polynomial, to improve the approximation $x^2 - 6x + 8$ for a quadratic factor by Bairstow's corrections.

		1	16.0	72.0	96.0	24.0
1	6.0	8.0	1	10.0	4.0	-8.0
	0.267	-0.133	1	4.0	-4.0	-8.0
				-28.0	-32.0	$\Delta = 240.0$
1	6.267	7.867	1	9.733	3.1363	-0.2247
	0.01546	0.05319	1	3.466	-4.7307	-0.6733
				-26.4521	-27.2670	$\Delta = 219.644$
1	6.28246	7.92019	1	9.71754	3.02975	0.00095
						0.00380

(The actual factor is $x^2 - 6.28238x + 7.91985$.)

The four italicized entries at the second stage, for example, arise from dividing the quotient of the initial division once again by the divisor $x^2 - 6.267x + 7.867$. It is easy to prove that they furnish the approximate elements of $g_2(B)$ and (provided that we divide by the determinant Δ) of its reciprocal, thus:

$$\begin{bmatrix} -26.4521 & 3.466 \\ -27.2670 & -4.7307 \end{bmatrix}, \quad \text{reciprocal } \frac{1}{219.644} \begin{bmatrix} -4.7307 & -3.466 \\ -27.2670 & -26.4521 \end{bmatrix}.$$

The elements of the reciprocal operate on the errors $\{-0.2247 \quad -0.6733\}$ to give the corrections that are shown at the left, $\{0.01546 \quad 0.05319\}$. All that is necessary on the work-sheet is to enter the value of Δ ; the computer will easily see how the multipliers consort with the errors, and which one in each pair of multipliers receives the change of sign. A similar routine holds for every stage. The convergence, if the procedure is strictly carried out, is quadratic, in the sense that each successive vector of corrections involves only squares, products and higher powers of the elements of the preceding error-vector. However, just as in the Newton-Raphson case (Whittaker and Robinson, 1944, p. 91), when satisfactory approximation is being attained a modification is possible; the matrix operating on the corrections, itself tending all the time to $\{g_2(B)\}^{-1}$, is then so stable that it need not be re-calculated. In this way, sacrificing quadratic but securing a good

linear convergence, we use a constant matrix and constant Δ and thus save the second line of division at each stage. It is not necessary to exemplify numerically this very convenient alternative procedure.

The Newton-Raphson correction and Bairstow's corrections are based on errors in *final* remainders. When we investigate the corresponding adjustments derived from errors in *penultimate* remainders, we find that they exist and are effective, but are more complicated to apply.

For example, with the same work-sheet as in Example 1, the second approximation would be computed thus:

$$(24 \times 30.4375)/(13.6406 \times 30.4375 + 0.1290 \times 25.1875) = 1.74579.$$

Since the actual root is 1.745761, we see that the Newton-Raphson process, which gave 1.74576 for second approximation, is not likely to find a rival in the above.

However, just as there is a *modified* Newton-Raphson or Bairstow procedure based on constant multipliers, so there is a modified procedure based on r.p.r., which can be achieved in the following convenient way. Suppose we are trying for a linear factor the exact value of which is $x - b$. Let us multiply the dividend $f_n(x)$ by $x - c$, where c is so chosen that the penultimate quotient arising from the division of $(x - c)f_n(x)$ by $x - b$ shall vanish for $x = b$, that is, shall contain $x - b$ as a factor. A simple application of the remainder theorem then gives c as follows. Let the quotient which arises from the division of $f_n(x)$ by $x - b$ be once again divided by $x - b$, exactly as in the second stage of a Horner transformation or a Newton-Raphson process; then if d_{n-2} and d_{n-1} are the two last coefficients in the second quotient, $c = -d_{n-1}/d_{n-2}$. This is the theory; in practice our divisor is only an approximation to $x - b$, and therefore we shall have only an approximation to c .

Example 3.—The same problem as Example 1. We take the two last entries in the line showing the second division, and construct $c = 30.4375/25.1875 = 1.208$, or 1.2 approximately. We therefore multiply the Laguerre polynomial by $x - 1.2$, and proceed with r.p.r.

		1 17.2	91.2	182.4	139.2	28.8
1 1.75	1	15.45	64.1625	70.1156	16.4977	28.8
1 1.74570	1	15.45430	64.22143	70.28865	16.49710	28.8
1 1.745761						

A good convergence is visible; in fact the successive errors tend to a ratio of about 1:340. On the other hand, a rougher value of c , such as $c = 1.25$, would have produced good but not striking convergence.

It is to be noted that ordinary r.p.r., applied with this particular divisor to the Laguerre polynomial, would diverge. Multiplication by

$x-c$ prepares the polynomial for the application of r.p.r., removing divergence and ensuring good convergence. It is necessary that the value of c should be rather close to the ideal value; that is, that the approximation to the linear divisor should not be too rough.

The device can be extended to divisors of higher degree than the first. For example, in the case of a quadratic divisor there exists a multiplier x^2-cx+d such that the penultimate quotient, when $(x^2-cx+d)f_n(x)$ is divided by this divisor, contains the divisor as a factor. This condition determines c and d as solutions of two simultaneous linear equations with coefficients appearing very naturally in the line showing the second division on the work-sheet; but we shall go no further into detail here.

7. CONCLUDING OBSERVATIONS

Some remarks of a general nature may be appended. It was pointed out in the earlier paper that, so far as division by a linear factor is concerned, penultimate remaindering is not new; it is merely ordinary iteration, with a particular way of breaking up the polynomial equation into two parts. Equally, and in a dual respect, when the divisor is of degree $n-1$, the process is not new; for the coefficients of the divisor form the vector $b=\{1 \ b_1 \ b_2 \ . \ . \ . \ b_{n-1}\}$, and if A is the rational canonical matrix (in the sense of the present paper) corresponding to the *dividend* (not the divisor), then penultimate remaindering is just a somewhat disguised way of building the vector sequence b, Ab, A^2b, A^3b, \dots , and this is a well-known way of approximating to the dominant latent root of A . It may be contrasted with the formation of the vector-sequence $u, A'u, (A')^2u, \dots$, which is equivalent (*cf.* Fry, 1945, p. 104) to the old method of D. Bernoulli (Whittaker and Robinson, 1944, p. 98) of building up a sequence of values u_i satisfying the difference equation with constant coefficients,

$$u_{i+n} - a_1 u_{i+n-1} + a_2 u_{i+n-2} - \dots + (-)^n a_n u_i = 0,$$

to which the polynomial is auxiliary.

Most of the papers included in the list of references given below are concerned not with penultimate remaindering but with Lin's later method. E. H. Sealy (to whose work reference should have been made in our previous paper) derived criteria for convergence of r.p.r. in the case of linear and quadratic divisors, and used accelerative processes of the usual kind. B. Friedman proposes a method based on iterated division, but using quotients, not remainders; we shall examine elsewhere the trans-

formation of errors in his method. Luke and Ufford extend Lin's method in the direction of resolving $f_n(x)$ not into two but into three or more polynomial factors, and they tabulate formulæ useful for this purpose; naturally these formulæ are more complicated than the simple bilinear recurrences of synthetic division.

REFERENCES TO LITERATURE

- AITKEN, A. C., 1950. *Proc. Roy. Soc. Edin.*, A, LXIII, 174-191.
 BAIRSTOW, L., 1920. *Applied Aerodynamics*, Longmans, Green & Co., London.
 FRIEDMAN, B., 1949. *Commun. Pure Appl. Math.*, II, 195-208.
 FRY, T. C., 1945. *Quart. Appl. Math.*, III, 89-105.
 LIN, S. N., 1939. Thesis for Sc.D., Massachusetts Institute of Technology.
 —, 1941. *Journ. Math. Phys.*, XX, 231-241.
 —, 1943. *Journ. Math. Phys.*, XXI, 60-77.
 LUKE, Y. L., and UFFORD, D., 1951. *Journ. Math. Phys.*, XXX, 95-109.
 SEALY, E. H., 1940. Thesis for Ph.D., University of Edinburgh.
 WHITTAKER, E. T., and ROBINSON, G., 1944. *The Calculus of Observations*, 4th ed., Blackie & Sons, London and Glasgow.

(Issued separately August 23, 1952)

XXIV.—The Solution of a Functional Equation.* By **A. H. Read**, M.A., United College, University of St Andrews.
Communicated by Dr D. E. RUTHERFORD.

(MS. received September 25, 1951. Revised MS. received February 29, 1952.
 Read May 5, 1952)

SYNOPSIS

Analytic solutions of the functional equation $f[s, \phi\{g(s)\}] = \phi(s)$, in which $f(s, w)$ and $g(s)$ are given analytic functions and $\phi(s)$ is the unknown function, are investigated in the neighbourhood of points ζ such that $g(\zeta) = \zeta$. Conditions are established under which each solution $\phi(s)$ may be given as the limit of a sequence of functions $\phi_n(s)$, defined by the recurrence relation $\phi_{n+1}(s) = f[s, \phi_n\{g(s)\}]$, the function $\phi_1(s)$ being to a large extent arbitrary.

1. INTRODUCTION

THE solution of functional equations of the type

$$f[s, \phi\{g(s)\}] = \phi(s), \quad (1)$$

where $\phi(s)$ is the unknown function, is closely related to the theory of iteration. The iterates of a function $g(s)$ are the functions $g_n(s)$ defined by the relations

$$g_1(s) = g(s), \quad g_n(s) = g\{g_{n-1}(s)\}, \quad (n = 2, 3, \dots).$$

If this sequence of functions converges for a fixed value of s , and if $g(s)$ is continuous, the limit ζ clearly satisfies the equation $g(\zeta) = \zeta$. If ζ satisfies this equation it is called a double point of the function $g(s)$, and such double points are of particular importance in connection with functional equations of the type (1).

Solutions of (1) have been given in certain special cases. In particular Koenigs (1884) studies Schroeder's equation

$$\phi\{g(s)\} = s\phi(s), \quad (2)$$

where $g(s)$ is a given function and s a given constant. He shows that, under certain conditions,

$$\lim_{n \rightarrow \infty} [g_n(s) - \zeta]/s^n$$

exists in a region containing a double point ζ of $g(s)$, and represents an

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

analytic solution of (2). Fatou (1920) considers the more general equation

$$A(s)\phi\{g(s)\}=\phi(s), \quad (3)$$

in which $A(s)$ and $g(s)$ are given functions. He shows that, subject to certain conditions, the infinite product

$$A(s) \prod_{r=1}^{\infty} A\{g_r(s)\}$$

converges and represents an analytic solution of the equation (3) in the neighbourhood of the origin.

These solutions exemplify a generalized iterative process which can be applied to equation (1). Let the functions $\phi_n(s)$ be defined by the recurrence relation

$$\phi_{r+1}(s)=f[s, \phi_r\{g(s)\}], \quad (4)$$

where $f(s, w)$ is a regular function of the two variables s and w . Then if, when s lies in some domain which contains also $g(s)$, $\phi_n(s)$ tends to a regular limit function $\phi(s)$ as $n \rightarrow \infty$, $\phi(s)$ will be a solution of the equation (1). It is the object of the present paper to establish conditions for the convergence of the sequence $\phi_n(s)$. The principal result is formulated in the theorem of § 4. The author wishes to express his indebtedness to the referees for a number of simplifications in the argument.

2. PRELIMINARY LEMMAS

LEMMA 1.—If $|k| < 1$, $c_n = u_{n+1} - ku_n$ and $c_n \rightarrow c$ as $n \rightarrow \infty$, then u_n tends to a limit as $n \rightarrow \infty$.

To prove this, write $v_n = u_n - c/(1-k)$. Then the equation $c_n = u_{n+1} - ku_n$ reduces to

$$v_{n+1} - kv_n = c_n - c.$$

Suppose that $\epsilon > 0$. Choose N so that $|c_n - c| < \epsilon$ when $n > N$. Then

$$v_n = v_0 k^n + \left[\sum_{r=0}^{N-1} + \sum_{r=N}^{n-1} \right] (c_r - c) k^{n-r-1},$$

and therefore

$$\overline{\lim}_{n \rightarrow \infty} |v_n| < \epsilon \overline{\lim}_{n \rightarrow \infty} \sum_{r=N}^{n-1} |k|^{n-r-1} < \frac{\epsilon}{1-|k|}.$$

It follows that $v_n \rightarrow 0$, and so $u_n \rightarrow c/(1-k)$, as $n \rightarrow \infty$.

LEMMA 2.—If $|a_r| < M/R^r$, $|b_r| < K/\rho^r$, for constants M , K , R and ρ , and the series $\sum_{n=1}^{\infty} a_n \left(\sum_{r=1}^{\infty} b_r s^r \right)^n$ is rearranged as the series $\sum_{n=1}^{\infty} c_n s^n$, then

$$|c_n| < \frac{MK}{R+K} \left(\frac{R+K}{R\rho} \right)^n.$$

(It is well known that this rearrangement does not alter the sum of the series provided that $|s|$ is sufficiently small.)

To prove this lemma we observe that $|c_n|$ does not exceed the coefficient of s^n in the rearrangement, as a power series in s , of the repeated series

$$\sum_{n=1}^{\infty} \frac{M}{R^n} \left(\sum_{r=1}^{\infty} \frac{K}{\rho^r} s^r \right)^n.$$

This repeated series is easily found to have the sum

$$\frac{MKs}{R\rho - (R+K)s},$$

and, by expanding this expression in powers of s , the required coefficient is

$$\frac{MK}{R+K} \left(\frac{R+K}{R\rho} \right)^n.$$

3. THE EQUATION $f[s, \phi(s)] = \phi(s)$

It is convenient to discuss the equation

$$f[s, \phi(s)] = \phi(s) \quad (5)$$

before considering the more general type (1). We shall write $|s| = \sigma$, and we shall suppose that $\sigma > 1$. Suppose also that the equation $f(0, w) = w$ has a solution β , and that $f(s, w)$ is a regular function of the two variables (s, w) in the vicinity of the place $(0, \beta)$. We may then write

$$f(s, w) = \sum_{r, i=0}^{\infty} a_{ri} s^r (w - \beta)^i. \quad (6)$$

We assume in this section that $\sigma > |a_{01}|$.

Let $\phi_1(s)$ be any function which is regular at the origin, and takes the value β there; and define the functions $\phi_n(s)$ successively by

$$\phi_{n+1}(s) = f[s, \phi_n(s)]. \quad (7)$$

In this section we shall show that, for each value of s in a neighbourhood of the origin, $\phi_n(s)$ tends to a limit as $n \rightarrow \infty$, and that this limit is a regular

function of s . We first show that, when $\phi_n(s)$ is expanded in powers of s , each of the coefficients in the expansion tends to a limit as $n \rightarrow \infty$.

Since $\phi_1(s)$ is regular at the origin, all the succeeding functions $\phi_n(s)$ are regular there, and we may write

$$\phi_n(s) = \sum_{p=0}^{\infty} a_p(n) s^p.$$

From the fact that $\phi_1(0) = \beta$, combined with (7), it follows that $\phi_n(0) = \beta$ for all n ; that is to say, $a_0(n) = \beta$ for all n . Therefore, from (6) and (7),

$$\phi_{n+1}(s) = \sum_{r,i=0}^{\infty} a_{ri} s^r \left(\sum_{p=1}^{\infty} a_p(n) s^p \right)^i. \quad (8)$$

The triple series on the right is absolutely convergent for sufficiently small values of s , and can therefore be rearranged in ascending powers of s . Thus, equating coefficients of s^m ,

$$s^m a_m(n+1) = a_{01} a_m(n) + P_n, \quad (9)$$

where P_n is a polynomial in those $a_r(n)$ for which $r < m$. The coefficients in P_n are independent of n .

We now make the induction hypothesis that, for each value of r less than m , $a_r(n)$ tends to a limit a_r as $n \rightarrow \infty$. This hypothesis is verified in the case $m=1$, since we have shown that $a_0(n) = \beta$ for all n . From the hypothesis it follows that P_n tends to a limit as $n \rightarrow \infty$. We now deduce from (9), by putting $c_n = P_n/s^m$, $k = a_{01}/s^m$, $u_n = a_m(n)$ in Lemma 1, that $a_m(n)$ tends to a limit as $n \rightarrow \infty$.

We next show that there exist numbers λ and ρ such that $|a_r(n)| < \lambda \rho^{-r}$ for all n . Let M be the maximum value of $|f(s, w) - \beta|$ when s, w are within circles $|s| < R$, $|w - \beta| < T$. Then, by the extension of Cauchy's inequalities to analytic functions of two variables,

$$|a_{r1}| < MR^{-r} T^{-1}. \quad (10)$$

Also, if M_T denotes the maximum of $|f(0, w) - \beta|$ on $|w| = T$, then

$$|a_{01}| = \lim_{T \rightarrow 0} M_T/T.$$

Hence, given $\epsilon > 0$, there exists η such that $|M_T/T - |a_{01}|| < \epsilon/2$ if $T < \eta$. Having chosen a value of T less than η , we can find R such that $|M/T - M_T/T| < \epsilon/2$. Then

$$|M/T - |a_{01}|| < |M/T - M_T/T| + |M_T/T - |a_{01}|| < \epsilon.$$

Since $\sigma > |a_{01}|$, it is therefore possible to choose R and T so that

$$M/T < \sigma. \quad (11)$$

Let

$$\lambda = T(\sigma - 1), \quad (12)$$

and choose ρ so that

$$\rho < R\sigma, \quad (13)$$

$$\frac{M\rho}{\lambda R\sigma} + \frac{MR}{T(R\sigma - \rho)} < 1 \quad (14)$$

(this being possible by (11) since the left-hand side of (14) tends to $M/T\sigma$ as $\rho \rightarrow 0$), and, for all $r > 0$,

$$\lambda\rho^{-r} > |a_r(1)|. \quad (15)$$

Since $\sum a_r(1)s^r$ is a power series with non-zero radius of convergence, this last condition can always be satisfied by suitable choice of ρ . Moreover, if we take for $\phi_1(s)$ the simplest permissible function, namely the constant β , it can be omitted altogether.

We now make the induction hypothesis that

$$|\alpha_r(\rho)| < \lambda\rho^{-r}$$

for all $r > 0$ and $\rho < n$. The hypothesis is satisfied in the case $n=1$ in virtue of (15). By Lemma 2 the coefficient of s^{m-r} in

$$\sum_{i=0}^{\infty} a_{ri} \left(\sum_{m=1}^{\infty} a_m(n)s^m \right)^i$$

has modulus less than

$$\frac{M\lambda}{R^r(T+\lambda)} \left(\frac{T+\lambda}{T\rho} \right)^{m-r}$$

provided that $m > r$; and, using (12), this expression may be written in the form

$$\frac{M\lambda}{R^r T \sigma} \left(\frac{\sigma}{\rho} \right)^{m-r}.$$

Therefore, picking out the coefficient of s^m in (8), we obtain

$$\begin{aligned} \sigma^m |a_m(n+1)| &< |a_{m0}| + \sum_{r=0}^{m-1} \frac{M\lambda}{R^r T \sigma} \left(\frac{\sigma}{\rho} \right)^{m-r}, \\ |a_m(n+1)| &< \frac{M}{R^m \sigma^m} + \frac{MR\lambda}{T(R\sigma - \rho)\rho^m} \left[1 - \left(\frac{\rho}{R\sigma} \right)^m \right] \\ &< \frac{\lambda}{\rho^m} \left[\frac{M}{\lambda} \left(\frac{\rho}{R\sigma} \right)^m + \frac{MR}{T(R\sigma - \rho)} \right] \\ &< \frac{\lambda}{\rho^m} \left[\frac{M\rho}{\lambda R\sigma} + \frac{MR}{T(R\sigma - \rho)} \right] \\ &< \lambda/\rho^m, \end{aligned}$$

where in the last four lines we have used (10), (13) twice, and (14). Hence by induction $|a_r(n)| < \lambda \rho^{-r}$ for all n and $r > 0$. Therefore

$$|a_r(n)s^r| < |\lambda s^r \rho^{-r}| = M_r, \text{ say,}$$

where M_r is independent of n and $\sum M_r$ is convergent if $|s| < \rho$. Hence on account of uniform convergence with respect to n ,

$$\lim_{n \rightarrow \infty} \sum_r a_r(n)s^r = \sum_r a_r s^r, \quad (|s| < \rho),$$

since each term of the series on the right is the limit, as $n \rightarrow \infty$, of the corresponding term on the left. Consequently $\lim_{n \rightarrow \infty} \phi_n(s)$ exists and is a regular solution of the functional equation (5) in the circle $|s| < \rho$.

That this solution is unique can be seen as follows. If $\phi(s)$ satisfies (5), then $\phi(0) = \beta$ where $f(0, \beta) = \beta$. Therefore, if $\phi(s)$ is regular at the origin, we may write

$$\phi(s) = \beta + \sum_{r=1}^{\infty} a_r s^r,$$

and, using (6), we must have

$$\beta + \sum_{r=1}^{\infty} a_r s^r s^r = \sum_{r,i=0}^{\infty} a_{r+i} s^r \left(\sum_{m=1}^{\infty} a_m s^m \right)^i. \quad (16)$$

Equating coefficients of s^n we obtain

$$a_n(s^n - a_{01}) = H(a), \quad (17)$$

where $H(a)$ is a function of a_1, a_2, \dots, a_{n-1} . Since we are assuming that $\sigma > |a_{01}|$, we cannot have $s^n - a_{01} = 0$ for any value of n . Therefore the coefficients a_n are uniquely determined in succession by equations of the type (17). Accordingly there cannot be more than one solution $\phi(s)$ which is regular at the origin and takes the value β there.

It is desirable to know how far the condition $\sigma > |a_{01}|$ is necessary for the truth of the results established in this section. If we remove this condition it is no longer true that the coefficient of s^m in $\phi_n(s)$ must tend to a limit as $n \rightarrow \infty$. This is seen by taking a particular example: if we try to solve the equation

$$3\phi(s) - 1 - s = \phi(2s)$$

by taking $\phi_1(s)$ to be the constant $\frac{1}{2}$, we find that the coefficient of s in $\phi_n(s)$ tends to $-\infty$ as $n \rightarrow \infty$.

It may nevertheless be true that the limiting process converges if the function $\phi_1(s)$ is chosen with more care in such cases. It is not indeed difficult to show that $\phi_1(s)$ can always be chosen so that each of the

coefficients in $\phi_n(s)$ tends to a limit; but it is an open question whether the power series determined by the limiting values of these coefficients is convergent.

4. THE EQUATION $f[s, \phi\{g(s)\}] = \phi(s)$

We shall prove the following theorem:

Let ζ be a double point of the function $g(s)$ such that $g(s)$ is regular at ζ and $0 < |g'(\zeta)| < 1$; and let $w = \beta$ be a solution of the equation

$$f(\zeta, w) = w$$

such that (i) $f(s, w)$ is regular in a domain of the plane (ζ, β) , and

(ii) $\left| \frac{\partial f}{\partial w} \right| < \frac{1}{|g'(\zeta)|}$ when $s = \zeta, w = \beta$. Then the functional equation

$$f[s, \phi\{g(s)\}] = \phi(s) \quad (1)$$

has just one solution $\phi(s)$ which is regular at ζ and takes the value β there. This solution can be given as $\lim_{n \rightarrow \infty} \phi_n(s)$, where the functions $\phi_n(s)$ are defined by the recurrence relation

$$\phi_{n+1}(s) = f[s, \phi_n\{g(s)\}], \quad (18)$$

and $\phi_1(s)$ is an arbitrary function regular at ζ and taking the value β there.

We shall write $g'(\zeta) = 1/s$, where evidently $|s| > 1$. By the work of Koenigs referred to in § 1, there exists a function $h(s)$, which is regular at ζ with $h'(\zeta) \neq 0$, and which satisfies the equation

$$h\{g(s)\} = \frac{1}{s} h(s) \quad (19)$$

in a neighbourhood of ζ . For such a function we clearly have $h(\zeta) = 0$; and, since $h'(\zeta) \neq 0$, the inverse function of $h(s)$ possesses a branch $h_{-1}(s)$ which is regular at the origin and takes the value ζ there

The transformation

$$Z = \frac{1}{s} h(s), \quad s = h_{-1}(sZ) \quad (20)$$

gives a one-one mapping of a neighbourhood of ζ in the s -plane on a neighbourhood of the origin in the Z -plane. From (19),

$$h\{g(s)\} = Z, \quad g(s) = h_{-1}(Z). \quad (21)$$

Consider a pair of functions $\phi(s)$, $\Phi(Z)$ defined in the neighbourhood of ζ and 0 respectively, and so related that

$$\Phi(Z) = \phi\{g(s)\}. \quad (22)$$

Then $\Phi(Z) = \phi\{h_{-1}(Z)\}$, and therefore, using (20),

$$\Phi(sZ) = \phi(s). \quad (23)$$

Conversely, if the functions satisfy (23), then $\Phi(sZ) = \phi\{h_{-1}(sZ)\}$, whence we deduce (22). Thus the functional equations (1) and

$$f[h_{-1}(sZ), \Phi(Z)] = \Phi(sZ), \quad (24)$$

considered in the neighbourhood of ζ and the origin respectively, are equivalent: the existence of a solution to one implies the existence of a solution to the other, the solutions being related by (22) and (23).

Now let the functions $\Phi_n(Z)$ be defined by

$$\Phi_n(sZ) = \phi_n(s),$$

so that by (22) we have also $\Phi_n(Z) = \phi_n\{g(s)\}$. Then, from (18),

$$\Phi_{n+1}(sZ) = f[h_{-1}(sZ), \Phi_n(Z)].$$

But the function $f[h_{-1}(sZ), w]$ is a regular function of the two variables (Z, w) in a domain of the plane $(0, \beta)$, and it takes the value β at this place. Moreover, when this function is expanded in powers of Z and $(w - \beta)$, the coefficient of $(w - \beta)$ is

$$\left[\frac{\partial f(s, w)}{\partial w} \right]_{s=\zeta, w=\beta}$$

which we know to have modulus less than $|s|$. By the work of § 3 it follows that, in a neighbourhood of the origin, $\Phi_n(Z)$ tends to a limit $\Phi(Z)$; and that $\Phi(Z)$ is the unique regular solution of (24) which takes the value β at $Z=0$. Since $\phi_n(s) = \Phi_n(sZ)$ for all n , therefore $\phi_n(s)$ tends to a limit $\phi(s)$ which satisfies $\phi(s) = \Phi(sZ)$. Consequently, owing to the equivalence of (1) and (24) which we have already remarked, $\phi(s)$ is a solution of (1) which is regular at ζ and takes the value β there: and it is the only such solution. This completes the proof of the theorem.

5. REGION OF VALIDITY OF THE SOLUTION

The solution which we have obtained for equation (1) is valid in a domain containing the double point ζ of $g(s)$. It is difficult to give a precise specification of the extent of this domain in the general case. We shall assume here that $f(s, w)$ is a regular function of the two variables for

all finite values of w , when s lies in the *attractive domain* of ζ . This attractive domain is defined as follows. Let $g_n(s)$ denote the n th iterate of $g(s)$. A double point ζ of $g(s)$ for which $|g'(\zeta)| < 1$ is contained in a certain domain Γ , such that $\lim_{n \rightarrow \infty} g_n(s_0) = \zeta$ whenever s_0 is a point of Γ (Montel, 1927, Chap. 8). This domain is called the attractive domain of ζ , and ζ itself is called an attractive double point.

We shall show that the sequence of functions $\phi_n(s)$ converges to a limit function which is regular, when s lies in Γ . In fact, we know that there is a domain D_1 such that $\phi_n(s)$ tends to a regular limit when s is in D_1 . We define a set of domains D_n by saying that D_{m+1} is the domain which is mapped on D_m by the transformation $s \rightarrow g(s)$. Γ is then contained in the union of all the D 's. We make the induction hypothesis that $\phi_n(s)$ tends to a regular limit when s lies in D_m . Now let s be a point in D_{m+1} . We have

$$\phi_{n+1}(s) = f[s, \phi_n\{g(s)\}],$$

and the right-hand side tends to $f[s, \phi\{g(s)\}]$ as $n \rightarrow \infty$. But this is a regular function of s . Therefore, by induction, $\phi_n(s)$ tends to a regular limit function when s is in any of the domains D_n ; and does so, consequently, when s is in the attractive domain of ζ .

6. THE REPULSIVE DOUBLE POINT

A repulsive double point of $g(s)$ is a double point ζ such that $|g'(\zeta)| > 1$. If (i) ζ is a repulsive double point of $g(s)$, (ii) β is such that $f(\zeta, \beta) = \beta$, (iii) $f(s, w)$ is a regular function of the two variables (s, w) in the vicinity of (ζ, β) , (iv) $\left(\frac{\partial f}{\partial w}\right)_{\zeta, \beta} \neq 0$, then, by the implicit function theorem for functions of two variables, the equation

$$f[s, \phi\{g(s)\}] = \phi(s)$$

may be expressed on the form

$$\phi\{g(s)\} = F[s, \phi(s)],$$

where $F(s, w)$ is an analytic function of the two variables (s, w) in the vicinity of (ζ, β) , and β is such that $\beta = F(\zeta, \beta)$. But, since $|g'(\zeta)| > 1$, there exists a branch $g_{-1}(s)$ of the inverse function which is regular at ζ and such that $|g'_{-1}(\zeta)| < 1$. Therefore the functional equation can be still further transformed to

$$\phi(s) = F[g_{-1}(s), \phi\{g_{-1}(s)\}],$$

which is of the form

$$\phi(z) = G[z, \phi\{g_{-1}(z)\}].$$

This is of the type previously treated, provided that $\left| \frac{\partial G}{\partial w} \right| < |g'(\zeta)|$ at (ζ, β) ; a condition which corresponds to $\left| \frac{\partial f}{\partial w} \right| > 1/|g'(\zeta)|$.

REFERENCES TO LITERATURE

- FATOU, P., 1920. "Sur les Équations Fonctionnelles", *Bull. Soc. Math. France*, XLVIII, 261.
 KOENIGS, G., 1884. "Recherches sur les Équations Fonctionnelles", *Ann. Sci. Éc. Norm. Sup. Paris*, ser. 3, 1, supplément, 19.
 MONTEL, P., 1927. *Leçons sur les Familles Normales de Fonctions Analytiques*, Paris.

(Issued separately August 23, 1952)

XXV.—The First Chemical Society, the First Chemical Journal, and the Chemical Revolution.* By James Kendall, M.A., D.Sc., LL.D., F.R.S., P.R.S.E. (With Two Plates.)

(Address of the President at a Meeting held on July 7, 1952)

(MS. received June 5, 1952)

1. THE FIRST CHEMICAL SOCIETY

IN very olden days chemists did not forgather merely as chemists; they merged themselves in broader organizations such as the Royal Society. The "chemical revolution", which had its real beginning with the work of Joseph Black and which culminated in the overthrow of the phlogiston theory by Lavoisier, aroused for the first time a popular interest in the special science of Chemistry. Until recently, world priority among the chemical societies that resulted therefrom was by general agreement conceded to the Chemical Society of Philadelphia, founded by James Woodhouse in 1792. The distinguished chemical historian Edgar F. Smith (1), late Provost of the University of Pennsylvania, may be quoted in this connection:

This was the first chemical society in the world. As far as can be learned, Woodhouse was its first and only president. This society lived about seventeen years. Its members favoured Lavoisier's doctrine of combustion. The minutes of the society have never been found, although diligent search has been made for them.

Copies of a number of publications of the society, however, have been preserved. One of them, Robert Hare's memoir on his invention of the oxyhydrogen blowpipe, represents a real landmark in scientific discovery.

The title of the Chemical Society of Philadelphia was not questioned until 1935. In April of that year, the Edinburgh University Chemical Society met to celebrate its Diamond Jubilee, being under the impression that it was founded only sixty years previously, and I was requested to prepare a speech for the occasion. Reading Sir William Ramsay's biography of Joseph Black to obtain some historical material for this oration, I came across the statement (2) that, among the correspondence of Joseph Black, Ramsay had discovered a sheet of paper, of which only the

* This address was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

date, 1785, was in Black's handwriting, entitled "List of the Members of the Chemical Society". Ramsay remarked regarding this as follows:

This may have been a society of persons residing in Edinburgh interested in Chemistry, but is more likely to have been a general society. . . . The only name that I can recognize is that of Dr Thos. Beddoes [the founder of the Pneumatic Institute and the "discoverer" of Sir Humphry Davy]; the names themselves would indicate that their possessors belonged to all parts of the kingdom.

The complete list of names—fifty-nine in all—is reproduced below.

John Webster	Hen. N. Ward	J. Alderson
Wm. Scott	Morgan Deasy	J. McElwaine
— Halliday	H. Pache	T. Gill
Jas. Forster	Jno. Gay	T. Willson
Sam. Black	Tho. McMorran	Frs. Montgomery
Jno. Black	Adam Gillespy	Thos. Swainson
Jas. Plumbe	Thos. Burnside	Archd. Webb
Wm. Johnston	Corn. Pyne	J. Crumbie
Thos. Clothier	J. Unthank	Geo. Marjoribanks
Henry Johnston	J. Barrow	T. Grieg
Peter Gernon	J. Donavan	J. Parr
Robt. Ross	Saml. Macoy	Alex. Stevens
L. van Meurs	G. Tower	Wm. Symonds
(amteemann batavis)	J. Sedgwick	Thos. Beddoes
Hugh Brown	T. Skeete	J. Thompson
John Boyton	Wm. Robertson	G. Kirkaldie
Edw. Fairclough	J. Sprole	J. Carmichel
Bicker McDonald	Thos. Cooke	Nich. Elcock
E. Galley	Thos. Edgar	Richd. Gray
An. Mann	Guyton Jolly	J. Hayle

The idea struck me that an examination of the register of students at the University of Edinburgh at that period might identify more of the names in the above list. This register was accordingly consulted, and within a quarter of an hour it was established that no fewer than fifty-three out of the fifty-nine were students attending Black's class in Chemistry at Edinburgh University during the years 1783–87. Thanks to the kindly zeal of Dr Alexander Morgan, five of the missing six were also subsequently traced as registered students of Joseph Black between 1780 and 1788. The only name on the list definitely unlocated is that of Peter Gernon.

In the light of the above discoveries, it became my pleasant duty to inform my fellow-members of the Edinburgh University Chemical Society at their "Diamond Jubilee Luncheon" that they had been called together under false pretences, since they were actually celebrating their Sesquicentenary! For it is quite plain that the society of 1785 was not, as Ramsay surmised, a *general* society. It was, on the contrary, a society consisting of those members of Black's class with a special interest in

chemistry—in other words, it was the Chemical Society of the University of Edinburgh. How long it survived after 1785 we have still no knowledge, but once the origin of its membership had been fixed, the possibility, noted by Ramsay, of locating “some one of their descendants in possession of some record of its proceedings and history” clearly became much more likely of realization.

Nineteen out of the fifty-nine members are on the list of medical graduates of the University of Edinburgh between the years 1784–90. This list also gives the nationality in each case, and it is interesting to find that, of the nineteen, only three were native Scots, three English, and the residual thirteen all Irish! Joseph Black was himself, of course, of Belfast ancestry, and the proportion of Irish students in the Scottish Universities has always been significant; but such a preponderance as this is very surprising. Whether its Irish element was responsible for the disruption of the society (it is ominous that Bicker McDonald and J. Unthank were both Hibernians) is a question that, perhaps just as well, we are not in a position to press.

The first chemical society in the world to complete a hundred years of continuous existence was the Chemical Society of London, the official records of which state that “on the 23rd of February, 1841, twenty-five gentlemen interested in the prosecution of Chemistry met together at the Society of Arts to consider whether it be expedient to form a Chemical Society”. Even after the publication (3) of my account of the origin of the Chemical Society of the University of Edinburgh, the Chemical Society of London could still, quite justifiably, object that no evidence had been adduced that the Chemical Society of the University of Edinburgh ever really functioned as a society. Nothing definite was known of its activities; there were no publications, no records of meetings, merely a list of names. How this deficiency was remedied will now be described.

2. THE FIRST CHEMICAL JOURNAL

In March 1947 I received a letter from the Rev. P. J. McLaughlin, D.Sc., of St Patrick's College, Maynooth, to the effect that he had discovered in the archives of the Royal Irish Academy, Dublin, a folio volume containing a collection of “Dissertations read before the Chemical Society instituted in the beginning of the Year 1785”. This volume, according to its title-page (Pl. I), had been presented to the Royal Irish Academy a century previously, in January 1846. No specific mention of its source was apparent anywhere in the volume, and it remained a puzzle to Dr McLaughlin until he was referred by chance by Dr Farrington,

Secretary of the Academy, to an article which I had written in *Endeavour* in 1942 on "Some Eighteenth-Century Chemical Societies" (4). A partial list of the contributors, which Dr McLaughlin sent me, consisted of names that all appear on Joseph Black's sheet.

I wrote at once to the Council of the Royal Irish Academy requesting them to be kind enough to loan the volume to the Royal Society of Edinburgh so that I might have the opportunity of inspecting it carefully. This favour was promptly accorded, and examination of its contents soon convinced me that I indeed held in my hands the first volume of the *Proceedings* of the Chemical Society of the University of Edinburgh. Presumably the unknown secretary of the society, to whose admirable diligence this record owes its being, was one of the formidable fraction of Irish members, and retained possession of his handiwork when he returned to his native country.

The book is a well-bound folio of 452 pages in copperplate manuscript, and contains thirty-two dissertations on topics of chemical interest. Some of the title-pages to the communications are true works of art (Pl. II). All of the contributors are on Joseph Black's list, with the exception of Nos. 22 and 27, Mr William Lecky and Mr S. Latham Mitchill, and it must be assumed that these gentlemen joined the society after its inception. At the conclusion of Paper 15 the words are added: Edin^r, 26th, Nov^r—1785; this is the only *direct* reference to Edinburgh in the volume. The beautifully clear handwriting throughout does not indicate, as I fondly supposed on first inspection, that students of that period were all experts in penmanship; its identity in numerous papers bears witness to the employment of professional scribes. These occasionally could not decipher chemical terms contained in the manuscript and left blanks for their subsequent insertion; where such insertions have been made they are still apt to be almost illegible.

Unfortunately, the thirty-two communications include very little matter of scientific importance, and my expectation that the volume might afford a valuable, as well as a unique, record of contemporary chemical thought has not been felicitously realized.* I had hoped to find the great Joseph Black himself participating in discussions upon the onslaught that Lavoisier was then launching against the phlogiston theory of Stahl, but Black suffered much from ill-health during his latter years and there is no note of his attendance at any of the meetings of the society. Refer-

* Perhaps I had pitched my standards too high, for I have since made a point of examining Volume I of the *Memoirs* of the Chemical Society of London, published in 1843, and although this contains communications by chemists of such eminence as Bunsen, Liebig, Graham and Playfair, there is not very much else in it that is really stimulating.

ences to "the learned Doctor" and to his work, however, are very frequent. Most of the papers are routine descriptive chemistry of the classical period, and one of the most promising of the few exceptions—"An Attempt to point out some of the Consequences which flow from Mr Cavendishes Discovery of the Component Parts of Water", by Mr Thomas Beddoes—peters out, after a tantalizing start, into an addendum entitled "A Conjecture concerning the Use of Manure"! Evidently the youthful Humphry Davy was not unduly disrespectful when he described Dr Beddoes, under whose direction he was working at the Pneumatic Institution at Clifton in 1799-1800, to be "as little fitted for a Mentor as a weather-cock for a compass".

This disappointment, however, emboldened me to make a second request to the Council of the Royal Irish Academy—a request which I should not have felt justified in making if the volume had proved to be of greater intrinsic value—namely that the Council should agree to return the volume to its original owners, the Chemical Society of the University of Edinburgh. To this request the Council of the Royal Irish Academy graciously assented, and I was accordingly able, at a meeting of the Society held on November 25, 1947, to make formal restoration to the first chemical society in the world of the first volume of its *Proceedings*, now lying on the lecture-table before me. Even if it has no great scientific significance, it does after all constitute the first journal extant of a purely chemical character, antedating by five years the *Annales de Chimie* initiated in Paris in 1790, and extracts from it may well prove worthy of wider publication.

Careful examination, indeed, has disclosed that one paper does contain matter of significant historical interest, as will be developed in the next section of my address. Meanwhile let me digress for a minute by reciting for your edification the introductory sentence of a topical communication by Mr Richard Webb "On Coal":

Among the innumerable Gifts which Nature has so liberally bestowed on Man, few seem to be more useful or better adapted to the conveniences of Life than the Lithanthrax or Pit Coal; and tho' not an exclusive, yet may be stiled a peculiar blessing to Britain, from their plenty, their excellence, and being found conveniently situated for exportation.

3. THE VIEWS OF MR JOHN CARMICHAELL ON PHLOGISTON

The penultimate paper of the first volume of the *Proceedings* of the Chemical Society of the University of Edinburgh, presumably * read in

* Since Papers 1 to 15 occupy from the beginning of the year 1785 to November 26, 1785 (see p. 349).

the latter part of the year 1786, is entitled, "Some Account of the Theories of Combustion, of Heat, of Light, and of Colour", by Mr John Carmichael. That this youthful philosopher was well aware that he had undertaken a heavy responsibility in discussing all these topics at once is seen by a Latin subscript to the title, "*Leni me corrigit manu*", which may be freely translated, "Don't spank me too severely for my mistakes!" We shall concern ourselves here solely with his views on the burning topic of phlogiston. How burning this question then was may be seen by the fact that it consumed five consecutive sessions of the Royal Society of Edinburgh in 1788, Sir James Hall appearing for the prosecution and Hutton for the defence. For the benefit of non-chemists, the essential point at issue may be very briefly (and, of course, very imperfectly) explained. The phlogistonists insisted that when a metal was heated in air and converted into an earth, or calx, phlogiston was expelled from it; the awkward fact that the calx weighed more than the original metal was explained by assigning to phlogiston a negative weight. The "pneumatic doctrine" of Lavoisier, on the other hand, maintained that the calx was a compound of the metal with the oxygen of the air. Mr Carmichael plunges right into the fight in his first paragraph:

Mr President: Stahl was the first who reduced Chemistry to a regular Science, and enriched it by addition, or at best the Extention, of Becher's doctrine of the Inflammable Earth, Phlogiston, or the Principle of Inflammability. Altho' perhaps we owe the advanced State in which we find the Science at this day to the introduction of this Principal, as it has all along afforded such easy and clear explanation of all Chemical Phenomena, yet it is frequently the cause of the utmost confusion and Ambiguity, as in speaking of it or endeavouring to describe it there are hardly two Persons who seem to possess any distinct or even the same ideas concerning it, or in describing it use the same language. It is covered with a dark Veil and protected by inexplicable Jargon.

After citing a number of examples of such jargon, Mr Carmichael proceeds:

However hypothetical this doctrine may appear, it has for long met with Courtesy, and Men of Science have unanimously countenanced it, and perhaps it might long have remained to be so, luted in the most perfect security, had not the Experimenting Genius of a Ray, Hales, Priestley, Bayen, Lavoisier or some other called in question on what Men of more confined or less exalted ideas of nature would have reckoned sacriligious to have attempted.

Before bidding adieu to the Manes of Phlogiston and wishing its Eternal Oblivion, the ingenuity of a Scheele and a Crawford demand our serious attention.

The body of Mr Carmichael's paper consists, in fact, of a discussion of the experiments and theories of "the ingenious Mr Scheele", "the studious Macquer", "the learned Lavoisier" and "the ingenious

Dr Lubbock". Having then explored the nature of heat, light and colour rather less diffusely, he concludes as follows:

And asking forgiveness of the Society for the many inaccuracies, perhaps neither few nor small, that I may have been led into, I conclude with the following line from Virgil which I think is applicable to this fanciful Proteus Phlogiston:

Venit summa dies—

Fuimus Troes, fuit Ilium et ingens gloria Teucrorum.

Now what is there remarkable in the fact that this young medical student should have announced so confidently in 1786 that the last hour of phlogiston had arrived? To answer this question I shall need to discuss the contemporary views of more famous chemists, particularly Joseph Black.

4. THE VIEWS OF JOSEPH BLACK ON PHLOGISTON

The opinion is still common among chemists that Joseph Black was among the last to abandon the theory of phlogiston and to accept the ideas of Lavoisier. P. Rousseau (5) has recently painted this picture:

Stahl's phlogiston loses its supporters one by one. . . . Chaptal has already banished phlogiston from his course of lectures at Montpellier; Berthollet follows suit in 1785. In 1786 it is the turn of Guyton de Morveau . . . the following year it is Fourcroy who burns what he has idolized . . . and claims the glory of being the first to teach officially Lavoisier's theory. Abroad, there is a much more lively resistance, especially in England and Germany. Slowly, however, the learned men renounce their errors; Jacquin in Holland, the famous Black. . . . There are only a few irreconcilables . . .

This misrepresentation originates essentially from John Robison, Professor of Natural Philosophy in the University of Edinburgh and the first Secretary of the Royal Society of Edinburgh, who edited the posthumous edition of Black's *Lectures on the Elements of Chemistry*. In the course of his editorial observations thereto, Robison (6) remarks:

Mr Lavoisier saw that his theory of combustion depended on the doctrine of latent heat, and was extremely anxious to obtain Dr Black's acquiescence. . . . Learning that Dr Black thought well of his theory, and had introduced it into his lectures, he wrote to him on July 14, 1790, as follows:

"J'apprends avec une joie inexprimable, que vous voulez bien attacher quelque merite aux idées que j'ai professé le premier contre la doctrine phlogistique. Plus confiant dans vos idées que dans les miennes propres, accoutumé à vous regarder comme mon maitre, j'étais en défiance contre moi même (*credat Judæus Apella* *) tant que je me suis écarté, sans votre aveu, de la route que vous avez si glorieusement suivie. Votre approbation,

* This sarcastic insertion is undoubtedly by Robison.

Monsieur, dissipe mes inquietudes, et me donne un nouveau courage. Je ne serai content jusqu'à ce que les circonstances me permettent de vous aller porter moi même le temoignage de mon admiration, et de me ranger au nombre de vos disciples. La revolution qui s'opère en France devait naturellement rendre inutile une partie de ceux attachés à l'ancien administration, il est possible que je jouisse du plaisir de la liberté, et le premier usage qui j'en ferai sera de voyager, et surtout en Angleterre, et à Edinburgh, pour vous y voir, pour vous entendre, et profiter de vos leçons et de vos conseils."

Dr Black wrote him a very plain, candid and unadorned letter in answer, expressing his acquiescence in his system. Mr Lavoisier answers this by praising in the highest terms the elegance of the style, the profoundness of the philosophy, etc. etc. and begs leave to insert the letter in the *Annales de Chymie*. Dr Black, who had been in very low spirits when he wrote that letter, and was much dissatisfied with its feebleness, was disgusted with what he now conceived to be artful flattery and refused to grant the request. Yet his letter appeared in that work before his refusal could reach Paris.

This wheedling, in order to screw out of Dr Black an acquiescence, on which he put a high value for the influence which it would have on the minds of others, was surely unworthy of Lavoisier. Dr Black was not only disgusted with the flattery, but seriously offended with its insincerity; and with a sort of insult on his common sense, by the supposition that he could be so wheedled, by a man whose publications never expressed the smallest deference for his opinions. For, by this time, Dr Black had read Mr Lavoisier's *Elements of Chemistry*, and the various dissertations by him and Mr de la Place, published in the *Memoirs of the Academy*. His name is not once mentioned, even in the dissertations on the measures of heat, where his doctrine of latent heat is delivered and employed as the result of Mr Lavoisier's own meditations. Nor is he named in those passages of the earlier dissertations, where the characters and properties of fixed air, and of the mild and caustic alkalis, are treated of. All appears to be the train of Mr Lavoisier's own thoughts, for which he was indebted to no man. Such inconsistency with the deference expressed in the above cited letters, provoked Dr Black to such a degree, that he resumed his critique on the nomenclature, and began to express his dissatisfaction with some parts of the theory, and his utter disapprobation of the unscientific and bullying manner in which the French chemists were trying to force their system on the world. But, by this time, his health had become so delicate, that the least intensity of study not only fatigued him, but made him seriously ill and forced him to give it up. I saw him but seldom at this time, being then in very bad health myself, but had this information from Dr Hutton, who shared all his thoughts.

Is there, in point of fact, an atom of truth in all the information that Robison gathered from Hutton (not an impartial witness; see p. 351), and in the imputations based thereon? That question can now be definitely answered in the negative, owing to a discovery made in 1949 by Dr McKie (7), and this section of my address can appropriately conclude with a condensed account of that discovery, since it leads me back once more to Mr Carnichaell.

In September 1793 two representatives of the Paris revolutionary committee called upon Lavoisier to search and seal his papers. Some letters written in English were taken away and sent to the Comité de

sûreté générale; they are now in the Archives de France, and among them are two from Joseph Black. The first, dated 24 October 1790, is the "very plain, candid, and unadorned letter" alluded to by Robison. The following extract is relevant here:

You have been informed that I endeavour in my Courses to make my Pupils understand the new principles and explanations of the Science of Chemistry which you have so happily invented and that I begin to recommend them as more simple & plain and better supported by Facts than the former system, and how could I do otherways? Your numerous & well contrived experiments have been performed with such uncommon accuracy & attention to every circumstance of any importance and with such quantities of the materials that nothing can be more satisfactory than the proofs of the Facts which you have investigated. And the System you have founded on these facts is so strictly connected with them and so simple & intelligible that it must be approved more & more every day and will even be adopted by many of those Chemists who have long been habituated to the former System: To gain them all is not to be expected, you know too well the power of habit which enslaves the minds of the bulk of mankind and makes them believe & reverence the greatest absurditys. I must confess that I felt the power of it myself, having been habituated 30 years to believe & teach the doctrine of Phlogiston as formerly understood. I felt much aversion to the new system which represented as an absurdity what I believed to be sound doctrine this aversion however which proceeded from the power of habit alone has gradually subsided, being over come by the clearness of your demonstrations & consistency of your Plan and tho there are still a few particulars which appear to be difficultys, I am satisfied that it is infinitely better supported than the former Doctrine; In this respect they cannot be compared, nor is this surprising when we consider that in composing yours you had the advantage of a multitude of new Experiments made with a degree of ingenuity & accuracy unknown to the Chemists of the former age. But tho the power of habit may prevent many of the older Chemists from approving of your Ideas, the younger ones will not be influenced by the same power; they will universally range themselves on your side of which we have experience in this university where the students enjoy the most perfect liberty of chuseing their philosophical opinions. They in general embrace your system and begin to make use of the new nomenclature in proof of which I send you two of their inaugural dissertations in which chemical subjects were chosen; these Dissertations are wrote entirely by the students; the Professors have no share in them. We read them before they are printed to see that there are no gross absurditys in them & give our advice if any are found: we sometimes find extravagant compliments to ourselves but have not the modesty or discretion to strike them out.

Black certainly does not seem to have written this letter in very low spirits, or to have any reason to be dissatisfied with its feebleness; Robison's subsequent statement that he was so disgusted by the artful flattery of Lavoisier's reply that he refused to permit its publication is refuted by a second letter, hitherto unpublished, sent from Edinburgh on December 28, 1790, which begins thus:

Mr Gahagan who is just returned to this Place made me happy with your letter of the 19th Nov^r. last & with the account he gave me of the ardor with which

you still pursue your philosophical researches. It gave me pleasure also to find that you are satisfied with the avowal I have made of my approbation of your System of Chemistry. You have my full consent to publish my letter. This consent I consider as a tribute I owe to truth and the eminent Rank you hold as a promoter & Patron of the Science of Chemistry: in publishing it you may leave out any parts of it which you think superfluous.

Lavoisier was, accordingly, perfectly justified when he published a French translation (by Madame Lavoisier) of Letter I in the *Annales de Chimie* in March 1791. As Dr McKie succinctly remarks: "The translation is unexceptionable and Black's wishes have been scrupulously observed. The whole incident does credit to two great chemists."

Even Robison admits that Black considered the death of Lavoisier (he was guillotined by the revolutionaries in May 1794) as "a great loss to science"; Black himself would probably have expressed his feelings much more emphatically. In 1796, when he was compelled at last to relinquish his long-famous lectures, he introduced Thomas Hope to his chemistry class in the following terms:

After having, for between thirty and forty years, believed and taught the chemical doctrines of Stahl, I have become a convert to the new views of chemical action; I subscribe to almost all M. Lavoisier's doctrines and scruple not to teach them. But they will be fully explained to you by my colleague and friend Dr Hope, who has had the advantage of having them from the mouth of their ingenious author.

It would be a mistake to interpret these words as indicating that Black was a recent convert to the new views. In my opinion, he was merely doing his best to give his young successor a flying start. Hope (8) is generally regarded as the first chemist to promulgate the doctrines of Lavoisier publicly in Great Britain, at the University of Glasgow in the winter of 1787-88, but I shall now proceed to cite strong presumptive evidence that Black actually anticipated him.

5. DR JOHN CARMICHAELL, PRESIDENT OF THE CHEMICAL SOCIETY OF EDINBURGH

By good fortune, Mr John Carmichaell is one of the nineteen original members of the Chemical Society instituted in 1785 whose name appears on the list of medical graduates of the University of Edinburgh (p. 348). This inspired me to secure a copy of his inaugural dissertation from the University Library, in the hope that it might contain matter of interest. My hope has, indeed, been richly fulfilled.

The dissertation, entitled *De Fermentatione*, was submitted in August 1787. The following extract from its introductory paragraphs will suffice

to prove that Mr Carmichaell was still of the same mind with regard to phlogiston as he was in 1786. I reproduce here the Latin text of the original, and append my own translation in a footnote:

Nostri ævi philosophi chemiam multum excoluere; hodierna experientia scientiæ multum lucis protulit; de multis processibus nunc fideliter disserere licet; æris necessitas nunc bene edocta; phlogiston, vagum principium, omnium fere consensu, ex chemia foras ejectum, dum umbræ loco, substantia doctis hodie foveatur.

His causis, doctrinam pneumaticam dictam elegi . . . *

On the basis of this doctrine, consequently, Dr Carmichaell develops his treatise on fermentation, with references to Black (*clarissimus*) and Lavoisier (*hodie primarius scientissimus*). His argument, however, is not very interesting from a modern point of view, for he is flogging what is now a very dead horse. Perhaps the most significant line in his thesis is on the title-page, where he proudly announces himself *Societ. Chem. Edin. Soc. Ext. et Pras. Annus.*† For this makes it clear, firstly, that the Society was not an ephemeral organization, but survived at least into the year 1787; and secondly, that the anti-phlogistic opinions expressed in Mr Carmichaell's paper to the Society (pp. 351-352) received such general approbation that his fellow-members almost immediately thereafter proceeded to elect him their president. It is nearly certain, moreover, in view of the final paragraphs of Black's first letter to Lavoisier (p. 354), that Mr Carmichaell's inaugural dissertation of 1787 was one of the two exhibits sent to Paris in 1790.

Although Black states that his students enjoyed "the most perfect liberty of chusing their philosophical opinions", few teachers will doubt that the majority moulded their views on the pronouncements of their professor, and the first volume of the *Proceedings* of the Chemical Society of the University of Edinburgh renders it practically certain that Black himself climbed over the phlogiston fence before 1785-86. Space considerations preclude the presentation of more than three short extracts. Mr William Scott, in the very first paper read before the Society, "On Chemical Attraction", observes:

Aqua regia dissolves gold. . . . Are we with Mr Scheele to suppose that here a double Elective Attraction takes place, that the Nitrous Acid attracts the

* The philosophers of our age have advanced chemistry considerably; current knowledge has brought much light on the science; it is now possible to discuss many operations exactly; the necessity of the air is now well inculcated; phlogiston, that mysterious principle, has by almost unanimous consent been cast out of chemistry, while in place of shadow, substance is to-day favoured by the well-informed.

For these reasons, I have chosen the doctrine called the "pneumatic doctrine".

† Extraordinary Member and President for the Year of the Chemical Society of Edinburgh.

phlogiston of the Muriatic Acid, and thus renders it capable of acting on the metal? No, for before his hypothesis be accepted he must prove the existence of phlogiston.

Mr Samuel Black, in Paper 3, "On Nitre", is still honestly bewildered:

Dr Stahl imagined that these phenomena were to be explained from the rapid union of the nitrous acid with Phlogiston. Mr Fourcroy on the contrary thinks that the appearances are to be explained from the rapid combustion of the Inflammable matter by means of the great quantity of pure Air disengaged in the process. Having stated the diversity of opinion among these illustrious chemists, it might well appear presumptuous in me to decide in favour of any of them.

But Mr William Haliday, discussing "The Calcination of Metals" in Paper 16, remarks:

Innumerable Hypotheses have been formed, a detail of which would be not only very tedious but also very useless, one of them however I cannot pass over in Silence, that of the illustrious Stahl, as it has been the one universally received till the late discoveries of Messieurs Bayen, Lavoisier, etc., which have thrown new light on this Subject.

These, and many similar quotations that might be added, make it apparent—to my mind at least—that Black, in 1785–86, while endeavouring to present both doctrines to his chemistry class as impartially as possible, was leading the more perspicacious among them from the old path to the new. When leisure allows me, I plan to look through *all* the Inaugural Dissertations of a chemical nature submitted to the Faculty of Medicine in the years between 1785 and 1790 to check this assumption and to discover how many other converts to Lavoisier's ideas, besides Dr Carmichaell, testified in Edinburgh during that period. The search will also, I hope, enable me to extend the "active life" of our first chemical society still further, perhaps even beyond 1790. Unfortunately, the dissertations are printed entirely in Latin, so that my task will not be an easy one.

I wish, in particular, to find out more about Dr John Carmichaell; his thesis does disclose a few additional details. First, the word *Scotus* after his name indicates that he was among the native minority of the original members of the Chemical Society (p. 348). Second, he was also a member of the Royal Medical Society, a student organization founded in 1737 (what pioneers Edinburgh students of that century were!) and still flourishing. Third, he dedicates his dissertation to his kinsman, Thomas Hay, Esquire, of the Royal College of Surgeons of Edinburgh. So far, my search for facts of his later life has been fruitless, but I am still

on his trail. Perhaps some day Volume II of the *Proceedings* of the Chemical Society of Edinburgh, issued under his presidency, will come to light and provide me with another of his pungent effusions to report upon to you. Meanwhile, for whatever errors I may have fallen into in my present commentary, *Leni me corrigit manu*.

Note added August 16, 1952.

Through the kindness of J. G. Kyd, Chairman of the Scots Ancestry Research Society, I have been furnished with the following data regarding Dr John Carmichael:

1. His father, Dr Michael Carmichael "of Hizellhead in the parish of Carmichall", was married on 25th April 1756 to Miss Mary Hay, daughter of John Hay, W.S., of the North-East parish of Edinburgh.

2. The Carmichael family appears to have been resident at Eastend, Carmichael (a village near Lanark), for centuries. John's elder brother Maurice became head of the family on the death of his uncle in 1789.

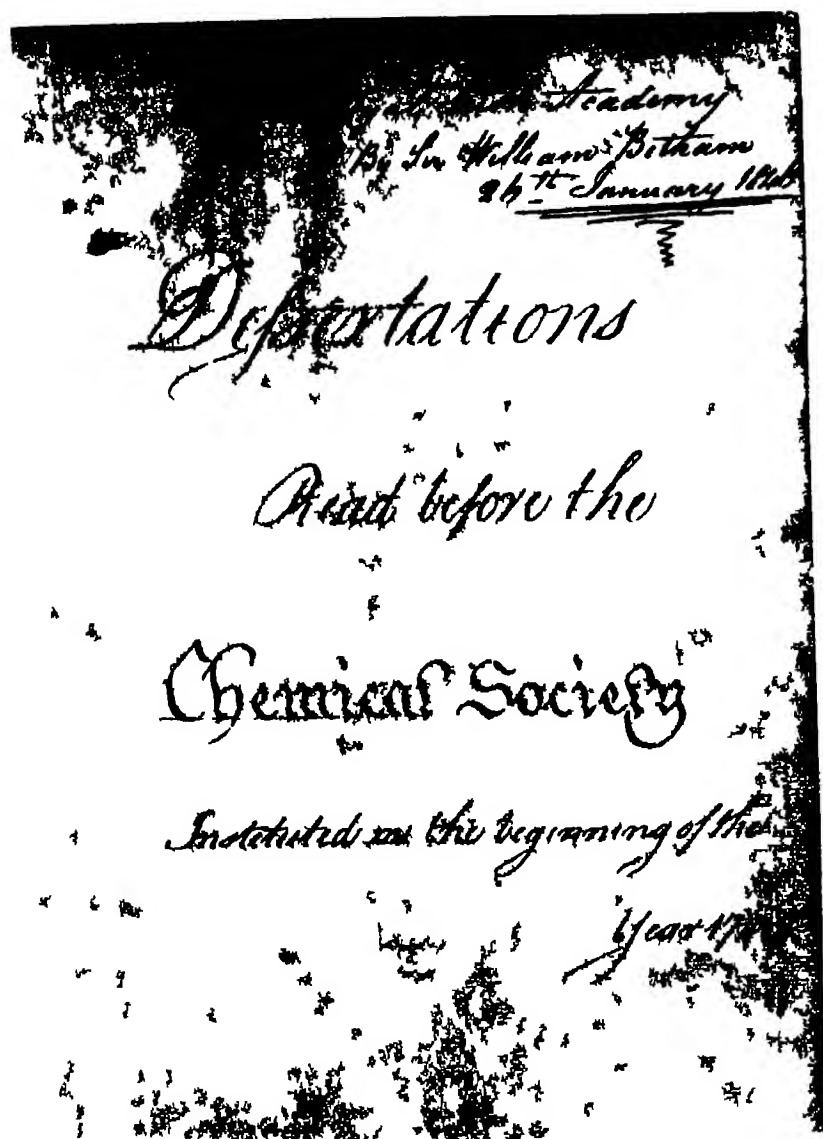
3. John Hay, W.S., a great-grandson of Sir John Hay of Barra, accompanied Prince Charles to France and was attainted in 1745. It was to his nephew Thomas Hay, M.D. of Edinburgh (born 1751—died 1816), the fifth son of Thomas, Lord Huntington, that Dr John Carmichael dedicated his inaugural dissertation in 1787.

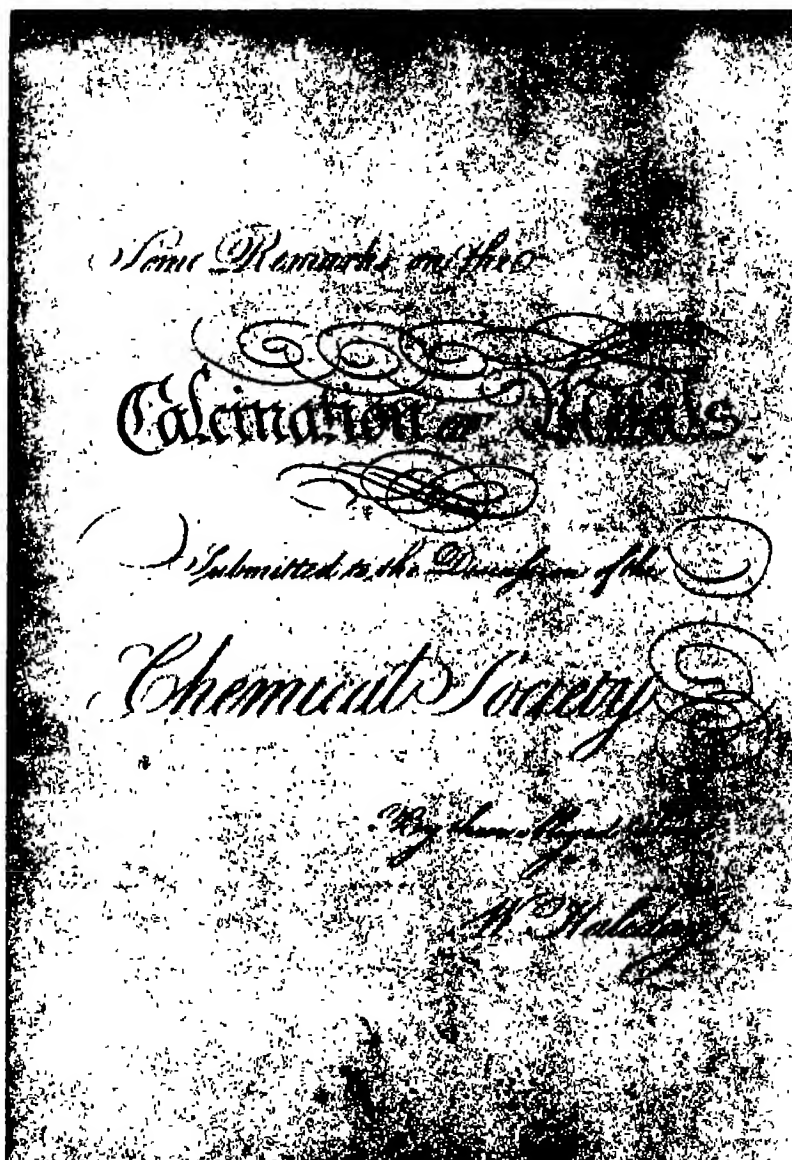
4. John himself was born at Hazelhead, Carmichael, on 20th February 1766. No details of his career after graduation have yet come to light.

REFERENCES TO LITERATURE

- (1) SMITH, 1914. *Chemistry in America*, D. Appleton & Co., 12.
- (2) RAMSAY, 1918. *The Life and Letters of Joseph Black, M.D.*, Constable & Co., 110.
- (3) KENDALL, 1935. *Journ. Chem. Educ.*, XII, 565.
- (4) KENDALL, 1942. *Endeavour*, I, 106.
- (5) ROUSSEAU, 1945. *Histoire de la Science*, Bronty, Fayard et Cie, 388.
- (6) ROBISON, 1803. *Lectures on the Elements of Chemistry, by the late Joseph Black*, II, 218-220.
- (7) MCKIE, 1949. *Notes and Records of the Royal Society of London*, VII, 1.
- (8) KENDALL, 1944. *Endeavour*, III, 120.

(Issued separately October 20, 1952)





XXVI.—The Normal Penetration of a Thin Elastic-Plastic Plate by a Right Circular Cone.* By J. W. Craggs, Ph.D., University College, Dundee. *Communicated by J. M. JACKSON, Ph.D.* (With Four Text-figures.)

(MS. received February 5, 1952. Revised MS. received May 28, 1952.
Read July 7, 1952)

SYNOPSIS

A study is made of the propagation of elastic and plastic deformation in a thin plate, initially unstressed, and of infinite extent, when it is penetrated normally by a cone moving with uniform velocity. The work is an extension of unpublished researches by Sir G. I. Taylor on the corresponding problem for a thin wire, and a summary of his results is included.

INTRODUCTION

THE propagation of plastic deformation in a thin wire, initially just taut and at rest, when it is struck by a projectile moving at right angles to it, has been studied by Professor Sir Geoffrey Taylor in unpublished work. In this paper a similar method is used for the discussion of the plastic and elastic waves set up in a thin plate, initially flat and free from stress, when it is penetrated normally by a conical-headed projectile moving with uniform velocity.

1. TENSION WAVES IN A THIN WIRE

The analogy between the behaviour of a thin plate and that of a thin wire is sufficiently close to make it worth while to summarise Taylor's results for the latter case.

Consider a thin wire of infinite length, initially taut, unstretched and at rest. Suppose a sharp-pointed projectile strikes the wire at $x=0$ and continues to move, perpendicularly to the wire, with uniform velocity W . Use ψ for the slope of the wire at a point x from the origin, t for the time, and ET for the tensile stress, where E is Young's modulus. Let τ be the ratio of the cross-sectional areas in the deformed and initial states, and ρ the density. It follows from similarity considerations that T , ψ , τ , U and V , where U , V are the longitudinal and transverse velocity components, are functions of $\lambda=x/t$ only.

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

Geometrical considerations lead to the relation

$$\frac{D\psi}{Dt} = -\cos \psi \left(\frac{\partial V}{\partial x} \cos \psi + \frac{\partial U}{\partial x} \sin \psi \right),$$

which reduces to

$$(\lambda - U)d\psi/d\lambda = \cos \psi \left(\cos \psi \frac{dV}{d\lambda} + \sin \psi \frac{dU}{d\lambda} \right). \quad (1)$$

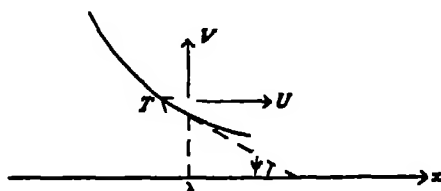


FIG. 1.—Variables used in wire problem.

The equations of motion

$$\rho \tau \sec \psi DU/Dt = E(d/dx)T\tau \cos \psi$$

and

$$\rho \tau \sec \psi DV/Dt = -E(d/dx)T\tau \sin \psi$$

reduce to

$$\rho \tau (\lambda - U)(dU/d\lambda) = -E(d/d\lambda)(\tau T \cos \psi) \quad (2)$$

and

$$\rho \tau (\lambda - U)(dV/d\lambda) = E(d/d\lambda)(\tau T \sin \psi). \quad (3)$$

The transverse velocity, V , may be eliminated between the three equations, and there results the equation

$$\{(\lambda - U)^2 - (ET/\rho) \cos^2 \psi\}(d\psi/d\lambda) = 0. \quad (4)$$

Equation (4) implies that $d\psi/d\lambda$ is zero except at the point $\lambda = \lambda_0$ corresponding to the velocity of transverse waves in a wire of tension $T = T_0$ (say). For $\lambda > \lambda_0$ and $\lambda < -\lambda_0$ the wire has zero slope. For $0 < \lambda < \lambda_0$ the slope is $\psi_0 = \tan^{-1} W/\lambda$, and for $-\lambda_0 < \lambda < 0$ the slope is $-\psi_0$. For a plastic-elastic wire there are three regions. For $\lambda - U > \sqrt{(E/\rho)}$ there is no disturbance. For $\sqrt{(E/\rho)} > \lambda - U > \sqrt{(TE/\rho)}$ the displacement is purely longitudinal, and for $\lambda - U < \sqrt{(TE/\rho)}$ the normal velocity is equal to the projectile velocity.

2. THE EQUATIONS OF MOTION FOR A THIN PLATE

For an isotropic thin plate deformed by a right circular cone moving along its own axis, which is normal to the plate, the deformation is symmetrical about that axis, and for an initially stationary unstressed

plate there is dynamic similarity, with a single independent variable $\lambda = r/t$, where r is the radius. Take radial and normal velocity components U and V ; radial and circumferential stress components $E\sigma_r$ and $E\sigma_\theta$; slope of the plate, measured in an axial plane, ψ , and thickness ratio τ (equal to current thickness : initial thickness). Then the geometrical relation (1) is still valid; the equations of motion reduce to

$$\cos \psi \frac{d}{d\lambda} (\lambda \tau \sigma_r \sin \psi) = - \frac{\lambda(\lambda - U)}{c^2} \tau \frac{dV}{d\lambda} \quad (5)$$

and

$$\cos \psi \frac{d}{d\lambda} (\lambda \tau \sigma_r \cos \psi) - \tau \sigma_\theta = - \frac{\lambda(\lambda - U)}{c^2} \tau \frac{dU}{d\lambda}, \quad (6)$$

where $c = \sqrt{E/\rho}$, and equation (4) is replaced by

$$\left\{ \sigma_r \cos^2 \psi - \left(\frac{\lambda - U}{c} \right)^2 \right\} \lambda \frac{d\psi}{d\lambda} + \sigma_\theta \sin \psi \cos \psi = 0. \quad (7)$$

3. THE ELASTIC SOLUTION

Consider first the case in which the plate is perfectly elastic and initially unstressed. Now the velocity of membrane-type waves of small amplitude is $\sqrt{T/\rho}$, where T is the radial tension in the plate. (This follows from (7), which shows that discontinuities of the derivative $d\psi/d\lambda$ occur only for the values of σ_r , which make the coefficient



FIG. 2.—The solution for a thin wire.

in double brackets equal to zero.) The first wave to travel outwards in the elastic plate must therefore be a tension wave travelling with velocity $c' = c/\sqrt{1 - \nu^2}$, appropriate to plane stress. Within the circle $\lambda = c'$ there will be an annulus in which the displacement v is zero. Now in elasticity theory changes in the thickness of the plate can be neglected; the equation of motion then becomes

$$\frac{d}{d\lambda} (\lambda \sigma_r) - \sigma_\theta = - \frac{\lambda(\lambda - U)}{c^2} \frac{dU}{d\lambda} \quad (8)$$

Next apply Hooke's law, with Poisson's ratio $\frac{1}{2}$. Then

$$u/(r - u) = \sigma_r - \frac{1}{2} \sigma_\theta \quad (9)$$

$$1/(1 - \partial u/\partial r) - 1 = \sigma_r - \frac{1}{2} \sigma_\theta \quad (10)$$

Now U , the velocity of an element, is given by $\delta r/\delta t$ with $r = u$ constant (cf. Taylor, 1948). In terms of $f = u/r$ this gives

$$U = -\lambda^2 f' / (1 - f - \lambda f'), \quad (11)$$

and when second-order terms are neglected the last four equations reduce to

$$\lambda^2(1 - 3\lambda^2/4c^2) \frac{d^2 f}{d\lambda^2} + 3\lambda(1 - \lambda^2/2c^2) \frac{df}{d\lambda} = 0. \quad (12)$$

Now u , and therefore f , is continuous at the singular point $\lambda = c'$, so the solutions of (12) are the one parameter family

$$f = -A \left\{ \frac{(1 - k^2 \lambda^2)^{1/2}}{2\lambda^2} - \frac{k^2}{2} \log \frac{1 + (1 - k^2 \lambda^2)^{1/2}}{k\lambda} \right\}, \quad (13)$$

where $k^2 = 1/c'^2$.

The corresponding values of the other variables are

$$U = -A(1 - k^2 \lambda^2)^{1/2} / \lambda, \quad (14)$$

$$\sigma_r = A \left\{ \frac{(1 - k^2 \lambda^2)^{1/2}}{3\lambda^2} + k^2 \log \frac{(1 - k^2 \lambda^2)^{1/2} + 1}{k\lambda} \right\}, \quad (15)$$

$$\sigma_\theta = A \left\{ -\frac{(1 - k^2 \lambda^2)^{1/2}}{3\lambda^2} + k^2 \log \frac{(1 - k^2 \lambda^2)^{1/2} + 1}{k\lambda} \right\}. \quad (16)$$

The table gives the numerical values of the variables, calculated from these equations.

λk	1.000	0.800	0.600	0.400	0.300	0.200	0.100	0.080	0.060	0.040
σ_r/A	0.000	0.637	1.102	1.891	2.730	4.881	15.09	21.86	37.28	80.99
σ_θ/A	0.000	0.403	0.546	0.459	0.081	-1.243	-10.60	-17.06	-32.04	-75.07
$-U/\lambda A$	0.000	0.703	1.667	4.293	7.950	18.35	77.04	116.8	208.0	468.4

3.1. The Transverse Wave-front

As the value of λ is reduced from c' , the radial tension σ_r increases steadily until the singularity of (7), with $\psi = 0$, is reached (this is the velocity of membrane-type waves).

At this value of λ the theory of differential equations leads us to consider the possibility of a discontinuity of ψ . Use the suffix zero for values in front of the wave (larger λ) and one for those behind. Then continuity of v and of mass leads to

$$(\lambda - U_1) \tan \psi_1 - (\lambda - U_0) \tan \psi_0 = V_1 - V_0, \quad (17)$$

$$(\lambda - U_1) \tau_1 \sec \psi_1 = (\lambda - U_0) \tau_0 \sec \psi_0. \quad (18)$$

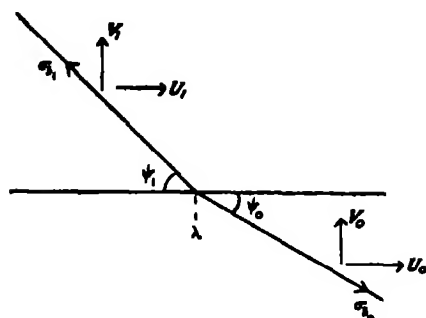


FIG. 3.—Conditions at a discontinuity.

The momentum equations give

$$\sigma_1 \tau_1 \sin \psi_1 - \sigma_0 \tau_0 \sin \psi_0 = \frac{(\lambda - U_1) \tau_1 (V_1 - V_0) \sec \psi_1}{c^2},$$

$$\sigma_1 \tau_1 \cos \psi_1 - \sigma_0 \tau_0 \cos \psi_0 = \frac{(\lambda - U_1) \tau_1 (U_0 - U_1) \sec \psi_1}{c^2}.$$

Elimination of V and τ_1/τ_0 leads now to

$$\frac{\tan \psi_1}{\lambda - U_1} \left\{ \left(\frac{\lambda - U_1}{c} \right)^2 - \sigma_1 \cos^2 \psi_1 \right\} = \frac{\tan \psi_0}{\lambda - U_0} \left\{ \left(\frac{\lambda - U_0}{c} \right)^2 - \sigma_0 \cos^2 \psi_0 \right\},$$

$$\frac{1}{\lambda - U_1} \left\{ \left(\frac{\lambda - U_1}{c} \right)^2 - \sigma_1 \cos^2 \psi_1 \right\} = \frac{1}{\lambda - U_0} \left\{ \left(\frac{\lambda - U_0}{c} \right)^2 - \sigma_0 \cos^2 \psi_0 \right\}.$$

A non-trivial solution of these equations is possible when

$$\sigma_0 \cos^2 \psi_0 = (\lambda - U_0)^2 / c^2, \quad (19)$$

and then

$$\sigma_1 \cos^2 \psi_1 = (\lambda - U_1)^2 / c^2. \quad (20)$$

3.2. The Bulged Region

Now consider the region behind the transverse wave. Then Hooke's law leads to

$$\frac{u}{r - u} = \frac{f}{1 - f} = \sigma_\theta - \frac{1}{2} \sigma_r \quad (21)$$

$$\frac{\sec \psi}{1 - \partial u / \partial r} - 1 = \sigma_r - \frac{1}{2} \sigma_\theta. \quad (22)$$

The dynamical equations, with constant thickness, reduce to

$$\cos \psi \frac{d}{d\lambda} (\lambda \sigma_s \sin \psi) = -\frac{\lambda(\lambda - U)}{c^2} \frac{dV}{d\lambda}, \quad (23)$$

$$\cos \psi \frac{d}{d\lambda} (\lambda \sigma_s \cos \psi) - \sigma_s = -\frac{\lambda(\lambda - U)}{c^2} \frac{dU}{d\lambda}, \quad (24)$$

and as before,

$$\left\{ \sigma_s \cos^2 \psi - \left(\frac{\lambda - U}{c} \right)^2 \right\} \lambda \frac{d\psi}{d\lambda} + \sigma_s \sin \psi \cos \psi = 0. \quad (25)$$

The required solution of these equations starts from the singularity

$$\sigma_s \cos^2 \psi - \left(\frac{\lambda - U}{c} \right)^2 = 0, \quad (26)$$

and a detailed examination of the singularity, omitted here for brevity, shows that there are no solutions through it unless σ_s is positive. The table of § 3, however, shows that σ_s changes sign for $\lambda/c' > 0.2$, so this condition implies $\sigma_s/E > 0.04$, and for most metals this is already outside of the elastic limit. Detailed examination of the elastic bulge is therefore omitted from this work.

3.3. *Transition to the Cone*

The considerations outlined in § 3.2 show that there will be no free elastic bulge in the plate. However, if the discontinuity in slope of the plate is at its point of contact with the cone, a different set of equations is needed, and such a transition may occur without any flat part of the plate being stressed beyond the elastic limit.

The equations of § 3.1, which were obtained by using only the conditions at a discontinuity of slope, still hold. Thus the transition to the conical region takes place when

$$\sigma_s = (\lambda - U_0)^2/c^2.$$

Now if the cone is assumed to be perfectly rough, the condition on the cone is $U=0$. Then (20) gives a relation between σ_s and ψ_1 . One of these parameters, say ψ_1 , may be chosen arbitrarily, then (17) gives the value of V_1 , which must be the velocity of the cone. The two parameters ψ_1 and A therefore correspond to the two-parameter system given by taking different cone angles and velocities.

If the cone is assumed to be perfectly smooth, the equation of motion

relative to the cone of the material in contact with it is

$$\frac{d}{d\lambda}(\lambda\sigma_\theta) - \sigma_\theta = -(\lambda^2/c^2) \sec^2 \psi_1 \frac{dU}{d\lambda}, \quad (27)$$

and use of the elastic equations derived from Hooke's law leads to

$$\lambda^2(1 - 3\lambda^2/4c^2) \frac{d^2 f}{d\lambda^2} + 3\lambda(1 - \lambda^2/2c^2) \frac{df}{d\lambda} + \frac{1}{2}\psi_1^2 = 0, \quad (28)$$

where ψ_1 is assumed small.

The solution of (28) is

$$U = -\lambda^2 f' = \frac{\psi_1^2 c^2}{3\lambda} \{1 - B(1 - k^2 \lambda^2)^{\frac{1}{2}}\}, \quad (29)$$

and the corresponding stresses are

$$\left. \begin{aligned} \sigma_r &= \frac{B\psi_1^2 c^2}{3} \left\{ \frac{(1 - k^2 \lambda^2)^{\frac{1}{2}}}{3\lambda^2} + k^2 \log \frac{(1 - k^2 \lambda^2)^{\frac{1}{2}} + 1}{k\lambda} \right\} - \frac{\psi_1^2 c^2}{9\lambda^2} + \frac{2\psi_1^2}{3} + D, \\ \sigma_\theta &= \frac{B\psi_1^2 c^2}{3} \left\{ -\frac{(1 - k^2 \lambda^2)^{\frac{1}{2}}}{3\lambda^2} + k^2 \log \frac{(1 - k^2 \lambda^2)^{\frac{1}{2}} + 1}{k\lambda} \right\} + \frac{\psi_1^2 c^2}{9\lambda^2} + D, \end{aligned} \right\} \quad (30)$$

where D and B are arbitrary.

The value of D may be found from consideration of the conditions at the transition to the cone. Determination of B depends on the condition at the tip of the cone. In fact, at that point $\lambda=0$, $U=0$ so $B=1$, and the stress components tend to infinity at the tip of the cone. This implies that the material becomes plastic before the tip of the cone is reached, and a new set of equations must be used. The difference between this case and the case to be considered in the next few sections is insignificant, and no special consideration will be given to it.

4. PLASTIC SOLUTIONS

When the chosen value of the parameter A is large (and this will be the case in most practical examples), the elastic limit of the material will be reached for a value of λ greater than that characteristic of transverse waves. There will then be a region in which the plate is flat and plastic.

The strain velocities in such a region are

$$\dot{\epsilon}_r = \frac{\partial U}{\partial r} = \frac{1}{t} \frac{dU}{d\lambda}, \quad \dot{\epsilon}_\theta = \frac{U}{r} = \frac{U}{t\lambda}, \quad \dot{\epsilon}_k = \frac{\lambda - U}{t} \frac{d}{d\lambda}(\log r). \quad (31)$$

The conservation of volume in plastic strain gives

$$\frac{dU}{d\lambda} + \frac{U}{\lambda} - \frac{\lambda - U}{\tau} \frac{d\tau}{d\lambda} = 0. \quad (32)$$

This equation is used only for the determination of the thickness.

The Huber-Mises flow equations reduce to

$$\frac{\dot{\epsilon}_s}{\dot{\epsilon}_\theta} = \frac{dU/d\lambda}{U/\lambda} = -\frac{2\sigma_s - \sigma_\theta}{\sigma_s - 2\sigma_\theta} = -K, \quad (33)$$

where K is introduced for convenience in the numerical work. The yield criterion (Mises-Hencky) is

$$\sigma_s^2 - \sigma_s\sigma_\theta + \sigma_\theta^2 = Y^2, \quad (34)$$

and strain hardening may be allowed for according to the law

$$Y = F(W), \quad (35)$$

where W is the plastic work done per unit volume. $F(W)$ is regarded as a known function.

Now W is given by

$$-(\lambda - U) \frac{dW}{d\lambda} = \sigma_s \frac{dU}{d\lambda} + \sigma_\theta \frac{U}{\lambda}. \quad (36)$$

Equations (32) to (36), with the dynamical equation, are sufficient for the determination of the dependent variables σ_s , σ_θ , τ , U , Y and W . They can be solved numerically, starting from the values given, for definite A , by the elastic solution at the least value of λ for which the yield condition is not violated.

For a perfectly plastic solution the yield stress Y is constant and equations (35) and (36) are not needed.

4.1. *The Bulge*

The solution given above holds, in general, for the annulus bounded internally by the circle corresponding to the velocity of transverse waves, which is given in § 3.1. Inside this circle the plate is bulged out of its own plane, and more general equations are required.

Now the rate of strain component $\dot{\epsilon}_s$ in such a bulge is

$$\dot{\epsilon}_s = \frac{1}{l} \frac{dU}{d\lambda} - \frac{\lambda - U}{l} \frac{d\psi}{d\lambda} \tan \psi, \quad (37)$$

so the conservation of volume equation is

$$\frac{dU}{d\lambda} + \frac{U}{\lambda} - \frac{\lambda - U}{\tau} \frac{d\tau}{d\lambda} - (\lambda - U) \frac{d\psi}{d\lambda} \tan \psi = 0. \quad (38)$$

The flow equation (33) is replaced by

$$\frac{\lambda(dU/d\lambda) - \lambda(\lambda - U)(d\psi/d\lambda) \tan \psi}{U} = -K, \quad (39)$$

and the rate of work equation by

$$-(\lambda - U) \frac{dW}{d\lambda} = \sigma_s \frac{dU}{d\lambda} - (\lambda - U) \sigma_s \frac{d\psi}{d\lambda} \tan \psi + \sigma_s U/\lambda. \quad (40)$$

Write

$$\sigma_s \cos^3 \psi - (\lambda - U)^2/c^2 = P, \quad (41)$$

then (7) becomes

$$\frac{d\psi}{d\lambda} + \frac{\sigma_s \sin \psi \cos \psi}{P\lambda} = 0. \quad (42)$$

Elimination of τ between (6) and (38) leads to

$$\left\{ \sigma_s \cos^3 \psi + \left(\frac{\lambda - U}{c} \right)^2 \right\} \lambda \frac{dU}{d\lambda} - 2(\lambda - U) \sigma_s \lambda \frac{d\psi}{d\lambda} \sin \psi \cos \psi \\ + \lambda(\lambda - U) \cos^3 \psi \frac{d\sigma_s}{d\lambda} - (\lambda - U) \sigma_s + \lambda \sigma_s \cos^3 \psi = 0,$$

and use of (39) to

$$\lambda(\lambda - U) \cos^3 \psi \frac{d\sigma_s}{d\lambda} - P\lambda \frac{dU}{d\lambda} - (\lambda - U) \sigma_s + \sigma_s(\lambda - 2KU) \cos^3 \psi = 0. \quad (43)$$

(40) reduces to

$$(\lambda - U) dW/d\lambda + (U/\lambda)(\sigma_s + K\sigma_s) = 0.$$

Now near the singularity $P=0$, (42) reduces to

$$\frac{dP}{d\psi} = - \frac{P\lambda}{\sigma_s \sin \psi \cos \psi} \left\{ \frac{\sigma_s}{\lambda} - \frac{3(\lambda - U)}{c^2} \right\}, \quad (44)$$

so for real solutions either ψ or σ_s must vanish at $P=0$. Moreover, the former possibility arises only for $\sigma_s > 0$,* and it is by no means clear that this can occur in practice. If it is assumed that $\sigma_s=0$, the value of σ_s in (20) follows from the yield condition, and (20) gives a relation between U_1 and ψ_1 . If one of these is chosen arbitrarily, the computation of the stresses and velocities in the bulged region is quite straightforward, apart from a rather tedious power series expansion to get away from the singularity.

* If it does, the initial value of $d\psi/d\lambda$ is the arbitrary parameter.

4.2. Conditions on the Cone

For a rough cone, the computation of the bulge is continued to a point at which U vanishes. The corresponding values of ψ and V give the angle and speed of the cone to which the solution corresponds.

For a smooth cone a trial-and-error solution must be used. At an arbitrary point of the bulge solution the equations are replaced by those appropriate to motion on a smooth cone, tangential to the plate at that point. This solution is then computed. If the choice of the junction between plate and cone is correct, then at $\lambda=0$ U vanishes. If this is not so, another cone solution must be computed, commencing from a different point of the bulge solution.

The equations on the cone are

$$\frac{d}{d\lambda}(\lambda\tau\sigma_s) - \tau\sigma_s = -\frac{\lambda(\lambda-U)}{c^2}\tau \sec^2 \psi_s \cdot \frac{dU}{d\lambda}, \quad (45)$$

where ψ_s is the cone angle, with

$$\frac{dU}{d\lambda} \cdot \frac{\lambda}{U} = -K, \quad (46)$$

$$\sigma_s^2 - \sigma_s\sigma_\theta + \sigma_\theta^2 = Y^2, \quad (47)$$

$$Y = F(W), \quad (48)$$

$$-(\lambda-U)(dW/d\lambda) = \sigma_s(dU/d\lambda) + \sigma_\theta(U/\lambda), \quad (49)$$

$$\frac{dU}{d\lambda} + \frac{U}{\lambda} - \frac{\lambda-U}{c} \frac{d\tau}{d\lambda} = 0. \quad (50)$$

A slight difficulty arises in the computation at points where K is infinite ($\sigma_s - 2\sigma_\theta = 0$). To avoid this trouble, (46) may be replaced by the Reuss equation,

$$\frac{dU/d\lambda - (d/d\lambda)(\sigma_s - \frac{1}{2}\sigma_\theta)}{U/\lambda - (d/d\lambda)(\sigma_s - \frac{1}{2}\sigma_\theta)} = -K, \quad (51)$$

for the computation near that point.

5. SUMMARY OF SOLUTION TYPES

For different velocities and angles of cone the solutions fall into four types.

(i) For low speeds and wide cone angles the plate is elastic and flat from infinity to the cone. There is a sharp bend where it meets the cone,

and a discontinuity of radial stress at this radius. On the cone there will, in general, be both elastic and plastic regions.

(ii) As the speed is increased or the cone angle reduced, a new solution type appears, in which there is a plastic flat region between the elastic part of the plane, and a bulged region. At the inside of the bulged region the plate is tangential to the cone, and for a smooth cone plastic deformation continues right to the tip of the cone.

(iii) If now the velocity of the cone is increased, the bulged region is reduced until at a certain velocity for any cone angle the bulge disappears, and the plate is flat except where it is in contact with the cone.

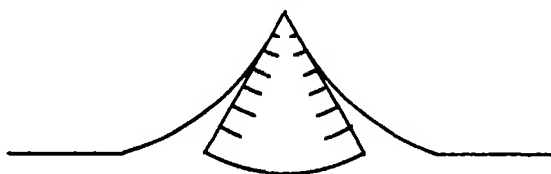


FIG. 4.—Form of the plate in the bulge solutions.

(iv) For a non-strain-hardening material, with a constant yield stress Y , or for any stress-strain curve with a horizontal asymptote, there is a definite limit $\sqrt{(Y/\rho)}$ to the possible transverse wave velocity. If then the cone has a large angle and a large velocity, the resolved part of the velocity of its surface in the plane of the plate will exceed this value. The above solutions then break down.

The solution may, however, be easily constructed. The plate will remain flat except where in contact with the cone, but at the circle of first contact with the cone there will be a shear stress, say ET , in the plate. This will be balanced by a pressure between the cone and the plate, concentrated over a small region near that circle.

The equations at the discontinuity of slope are then

$$\sigma_{\theta 1} \tau_1 \sin \psi_1 + \tau_1 T \cos \psi_1 = \{(\lambda - U_1) \tau_1 V_1 \sec \psi_1\} / c^2$$

and

$$\sigma_{\theta 1} \tau_1 \cos \psi_1 - T \tau_1 \sin \psi_1 - \sigma_{\theta 0} \tau_0 = \{(\lambda - U_1) \tau_1 (U_0 - U_1) \sec \psi_1\} / c^2,$$

where ψ_1 is the angle of the cone. The equations corresponding to (19) and (20) are then

$$(\lambda - U_1)^2 / c^2 - \sigma_{\theta 1} \cos^2 \psi_1 - T \cos^2 \psi_1 / \sin \psi_1 = 0$$

and

$$(\lambda - U_0)^2 / c^2 - \sigma_{\theta 0} - (T \cos \psi_1 \sin \psi_1) / (\lambda - U_1) = 0,$$

and there is no restriction on the value of λ at which such a discontinuity is possible.

ACKNOWLEDGMENT

My thanks are due to Professor Sir Geoffrey Taylor for suggesting the problem, and for his help and encouragement.

REFERENCE TO LITERATURE

TAYLOR, G. I., 1948. "The Formation and Enlargement of a Circular Hole in a Thin Plastic Sheet", *Quart. Journ. Mech. App. Math.*, 1, pp. 103-124.

(Issued separately October 20, 1952)

XXVII.—The Rotational Field behind a Bow Shock Wave in Axially Symmetric Flow using Relaxation Methods.*

By **A. R. Mitchell**, Ph.D., and **Francis McCall**, B.Sc.,
United College, University of St Andrews. *Communicated by*
Dr D. E. RUTHERFORD. (With Five Text-figures.)

(MS. received April 12, 1952. Read July 7, 1952)

SYNOPSIS

The relaxation technique of R. V. Southwell is developed to evaluate mixed subsonic-supersonic flow regions with axial symmetry, changes of entropy being taken into account. In the problem of a parallel supersonic flow of Mach number 1.8 impinging on a blunt-nosed axially symmetric obstacle, the new technique is used to determine the complete field downstream of the bow shock wave formed. Lines of constant vorticity and Mach number are shown in the field, and where possible a comparison is made with the corresponding 2-dimensional problem.

1. INTRODUCTION

IN supersonic flow past a symmetrically placed infinite cone, the flow behind the shock wave forms a conical field (Busemann, 1929) provided the shock surface is attached to the tip of the cone. The shock, which is a co-axial circular cone, is everywhere of constant strength and so the flow behind the shock wave is irrotational. The physical characteristics of such a flow have been theoretically determined for the possible range of incident Mach numbers and cone angles by Hantzsche and Wendt (1942) and Kopal (1947). If, however, the boundary conditions are such that an attached shock is impossible, the conical nature of the field is completely destroyed, and no theoretical solution to such a problem seems likely in the near future.

The attempts so far made to obtain a numerical solution of the field behind a bow shock wave in axially symmetric flow include those of Maccoll and Codd (1945) and Drougge (1948). In these investigations where the forward Mach number ranged from 1.50 to 2.15, relaxation methods were used to evaluate the subsonic region, and isentropic conditions were assumed downstream of the shock wave.

In the present paper, Southwell's relaxation technique is modified and

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

used to determine the complete mixed subsonic-supersonic rotational field downstream of the detached bow wave formed when a uniform parallel flow of Mach number 1.8 impinges on a square-nosed axially symmetric obstacle. At such an incident Mach number, an attached shock wave is only possible up to a semi-cone angle of 37° . The method of treatment is substantially similar to that employed in the corresponding 2-dimensional problem by one of the present authors (1951), and to avoid repetition only the main differences will be described in detail.

2. THE FUNDAMENTAL EQUATIONS

Cylindrical polar co-ordinates x, r are used, x being measured along the axis of symmetry and r perpendicular to it. The components of the velocity q in these two directions are u, v respectively. The stream function ψ is then defined by the equations

$$u = \frac{1}{\rho r} \frac{\partial \psi}{\partial r}, \quad v = -\frac{1}{\rho r} \frac{\partial \psi}{\partial x},$$

where ρ denotes the density.

The condition which ψ must satisfy for non-viscous, steady, compressible flow is (Vazsonyi, 1945)

$$\frac{\partial}{\partial x} \left(\frac{1}{\rho r} \frac{\partial \psi}{\partial x} \right) + \frac{\partial}{\partial r} \left(\frac{1}{\rho r} \frac{\partial \psi}{\partial r} \right) + \frac{r p}{R} \frac{\partial S}{\partial \psi} = 0, \quad (1)$$

where p, S denote the pressure and entropy respectively and R is the gas constant. This reduces to

$$\nabla^2(\chi\psi) - \psi(\nabla^2\chi) - \frac{\chi}{r} \frac{\partial \psi}{\partial r} + \frac{p r^2}{R \chi} \frac{\partial S}{\partial \psi} = 0, \quad (2)$$

where $\chi = \rho^{-1}$ and ∇^2 denotes $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial r^2}$.

The flow behind a shock wave is adiabatic, but not necessarily isentropic. The entropy, however, maintains a constant value along each streamline, and so Bernoulli's equation for the r th streamline behind the shock wave may be written

$$\left(\frac{q}{c_s} \right)^2 = \frac{2}{\gamma - 1} \left\{ 1 - \left(\frac{\rho}{(\rho_s)_r} \right)^{\gamma-1} \right\},$$

where c_s denotes the velocity of sound at a stagnation point, $(\rho_s)_r$ the value of the stagnation density on the r th streamline, and γ is the ratio of the specific heats.

In terms of χ and ψ , this equation becomes

$$\frac{2}{\gamma-1} \left\{ 1 - \left(\frac{\chi}{(\chi_s)_r} \right)^{2(1-\gamma)} \right\} = \frac{\chi^4}{r^2 \nu_r^2} \left\{ \left(\frac{\partial \psi}{\partial x} \right)^2 + \left(\frac{\partial \psi}{\partial r} \right)^2 \right\}. \quad (3)$$

Equations (2) and (3) are the fundamental equations for the axially symmetric field downstream of the shock wave.

For convenience, these two equations may be written in the non-dimensional form,

$$\nabla^2(\chi\psi) - \psi \nabla^2 \chi - \frac{\chi}{r} \frac{\partial \psi}{\partial r} + \frac{1}{\gamma \nu_r^2} \frac{\partial r^2}{\chi} \frac{1}{R} \frac{\partial S}{\partial \psi} = 0 \quad (4)$$

and

$$\frac{2}{\gamma-1} \left\{ 1 - \chi^{2(1-\gamma)} \right\} = \frac{\nu_r^2 \chi^4}{r^2} \left\{ \left(\frac{\partial \psi}{\partial x} \right)^2 + \left(\frac{\partial \psi}{\partial r} \right)^2 \right\}. \quad (5)$$

To obtain these equations, h , a significant length in the problem under consideration, and $2\pi G$, the mass flow per second under free stream conditions, have been introduced. With a slight change of notation,

χ , ψ , p , x , r now stand for the non-dimensional quantities $\frac{\chi}{(\chi_s)_r}$, $\frac{\psi}{G}$, $\frac{p}{(p_s)_r}$, $\frac{x}{h}$, $\frac{r}{h}$, where $(\chi_s)_r$ and $(p_s)_r$ are the stagnation values of χ and p on the r th streamline behind the shock, and the flow parameter on this streamline has been expressed as

$$\nu_r = \frac{G}{(\rho_s)_r \nu_s h^2}.$$

The pressure p can be eliminated by using the dimensional adiabatic law

$$\frac{p}{\rho^\gamma} = \frac{(p_s)_r}{(\rho_s)_r^\gamma},$$

and equation (4) becomes

$$\nabla^2(\chi\psi) - \psi \nabla^2 \chi - \frac{\chi}{r} \frac{\partial \psi}{\partial r} + \frac{1}{\gamma \nu_r^2} \frac{\partial r^2}{\chi^{2\gamma+1}} \frac{1}{R} \frac{\partial S}{\partial \psi} = 0 \quad (6)$$

The value of the flow parameter ν_r is given by one of the present authors (1951),

$$\frac{\nu_r^2}{\nu^2} = \left(\frac{p_2}{p_1} \right)^{\frac{2}{\gamma-1}} \left[\frac{(\gamma-1) \frac{p_2}{p_1} + (\gamma+1)}{(\gamma+1) \frac{p_2}{p_1} + (\gamma-1)} \right]^{\frac{2\gamma}{\gamma-1}}, \quad (7)$$

where ν is the flow parameter in front of the shock, and p_1 , p_2 denote the

values of the pressure upstream and downstream respectively of the shock wave on the r th streamline.

3. THE RELAXATION METHOD

The Southwell relaxation method (Green and Southwell, 1943) is used, and in all calculations the value 1.4 has been taken for γ .

A square network of non-dimensional mesh size a is used to cover the field, and the finite difference approximations to equations (5) and (6) are obtained as

$$R_0 \equiv \chi_0^{-4} - \chi_0^{-2(1+\gamma)} - \frac{(\gamma-1)(v_r)_0^2}{4a^2 r_0^2} \left[4\psi_0^2 + \sum_{i=1}^4 \psi_i(\psi_i - 2\psi_0) \right] \quad (8)$$

and

$$F_0 \equiv \sum_{i=1}^4 \chi_i(\psi_i - \psi_0) - \frac{\chi_0 a^2}{2r_0} (\psi_2 - \psi_0) + \frac{r_0^2 a^2}{\gamma(v_r)_0^2 \chi_0^{2(\gamma+1)}} \frac{1}{R} \left(\frac{\partial S}{\partial \psi} \right)_0, \quad (9)$$

where the suffixes 0, 1, 2, 3, 4 are as shown in fig. 1. Thus for a given

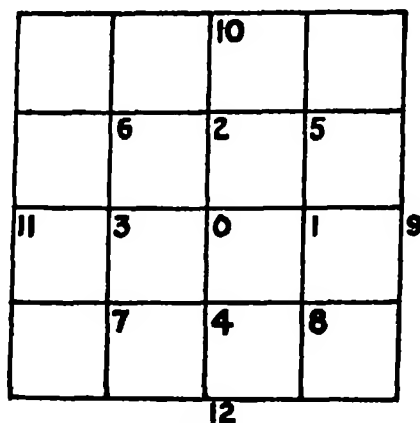


FIG. 1.

ψ -distribution, R_0 , χ_0 , and hence the residuals F_0 can be calculated for each node of the net, using equations (8) and (9). The ψ -distribution is now modified either by using a relaxation pattern, or by trial and error, in order to reduce the residuals. The changes in the residuals δF_j ($j=0, 1, 2, 3, 4$) following a change $\delta\psi_0$ in ψ_0 are given by the formulæ

$$\delta F_0 = - \left\{ \sum_{i=1}^4 [\chi_i + G_i(\psi_i - \psi_0)] + G_0 \left[\frac{a}{2r_0} (\psi_2 - \psi_0) + \frac{(2\gamma+1)r_0^2 a^2}{\gamma(v_r)_0^2 \chi_0^{2(\gamma+1)}} \frac{1}{R} \left(\frac{\partial S}{\partial \psi} \right)_0 \right] \left[4\psi_0 - \sum_{i=1}^4 \psi_i \right] \right\} \delta\psi_0 \quad (10)$$

and

$$\delta F_j = \left\{ \chi_0 + G_0(\psi_0 - \psi_1) \left(4\psi_0 - \sum_{i=1}^4 \psi_i \right) - \frac{(\gamma+1)r_j^2 a^2}{\gamma(v_r)_j^2 \chi_j^{\gamma+1}} G_j(\psi_0 - \psi_1) \frac{1}{R} \left(\frac{\partial S}{\partial \psi} \right)_j + Q_j \right\} \delta \psi_0, \quad (11)$$

where

$$G_0 = \frac{(\gamma-1)(v_r)_0^2}{2a^2 r_0^2} [f'(R)]_0,$$

$$G_j = \frac{(\gamma-1)(v_r)_j^2}{2a^2 r_j^2} [f'(R)]_j, \quad j=1, 2, 3, 4,$$

$$f'(R) = \frac{\chi^3}{2(\gamma+1)\chi^{3(1-\gamma)} - 4},$$

and where Q_j has the following values for $j=1, 2, 3, 4$:

$$Q_1 = -\frac{aG_1}{2r_1}(\psi_0 - \psi_1)(\psi_3 - \psi_4),$$

$$Q_2 = -\frac{aG_2}{2r_2}(\psi_0 - \psi_2)(\psi_{10} - \psi_4) + \frac{a\chi_2}{2r_2},$$

$$Q_3 = -\frac{aG_3}{2r_3}(\psi_0 - \psi_3)(\psi_5 - \psi_7),$$

$$Q_4 = -\frac{aG_4}{2r_4}(\psi_0 - \psi_4)(\psi_8 - \psi_{12}) - \frac{a\chi_4}{2r_4}.$$

In deriving these formulæ, it has been assumed that $(v_r)_0^2$ and $\left(\frac{\partial S}{\partial \psi}\right)_0$ do not vary with the change $\delta\psi_0$ in ψ_0 . The consequent inaccuracy in the relaxation pattern is unimportant.

4. THE PROBLEM

The problem under consideration is that of a uniform parallel air stream of Mach number 1.8 flowing past a square-nosed cylindrical obstacle. The position and shape of the bow shock wave formed in front of the obstacle was obtained from a photograph taken in the N.P.L. (Holder, North and Chinneck, 1949). The part of the field examined downstream of the shock wave extends from the axis of symmetry to the first streamline not appreciably deflected by the shock wave. At this streamline the shock wave makes an angle $\sin^{-1} \frac{1}{1.8}$ with the line of flow.

In this problem, the length h is chosen to be five times the semi-width of the obstacle, and $2\pi G$ to be the free stream flow per second through a circular cylinder of radius h . The first streamline not appreciably deflected by the shock wave is found to be a distance $2h$ from the axis of symmetry. The shock wave cuts this streamline in the point A , the sonic line in B and the axis of symmetry in C . The obstacle is bounded by DEF . The free stream conditions of this problem are $M=1.80$, $\rho=.287$, $\nu=1.262$, $R=.0324$, and $\psi=r^2$.

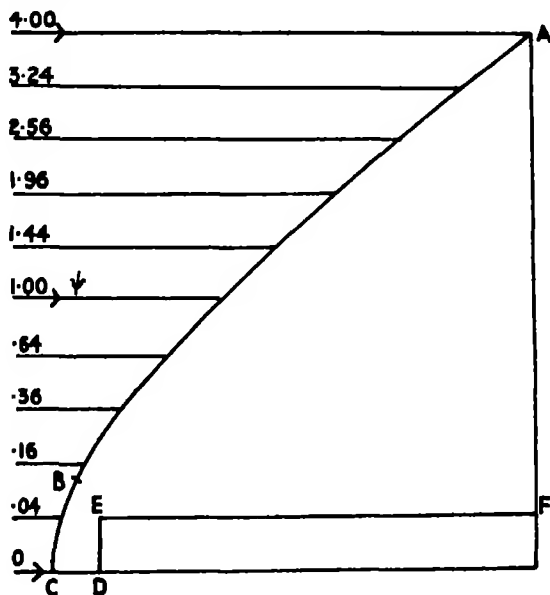


FIG. 2.

The field inside $ABCDEF$ (fig. 2), where the rotational flow is mixed subsonic and supersonic, is now evaluated by the methods of the previous section. A brief summary of the method used need only be given, as it is similar to that described fully by one of the present authors (1951) in the evaluation of the similar 2-dimensional problem. The boundary conditions again consist of a knowledge of the stream function ψ along $ABCDEF$ together with the Mach number M and the slope of the streamlines on the downstream side of the shock. In particular, knowledge of the pressure ratio $\frac{p_2}{p_1}$ across the shock enables the flow parameter ν , to be calculated from equation (7) for every streamline. The quantity $\frac{1}{R} \frac{\partial S}{\partial \psi}$ on the downstream side of the shock wave is found from a knowledge of

on the downstream side of the shock wave is found from a knowledge of

the entropy increase $\frac{1}{R}(S_r - S_0)$ across the shock wave at all points.

$\frac{\nu_r^2}{\nu^2}$ and $\frac{1}{R} \frac{\partial S}{\partial \psi}$ are each plotted against ψ in fig. 3. These quantities are constant along each streamline, and hence fig. 3 can be used everywhere downstream of the shock. As in the 2-dimensional problem, the ratio of the flow parameters on the two sides of the shock is as great as 1.5 for the streamline which crosses the shock normally.

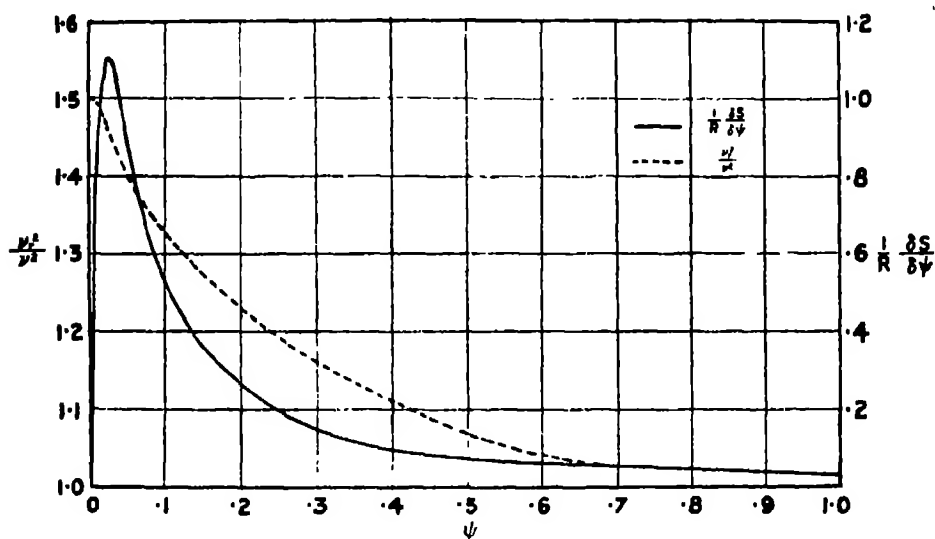
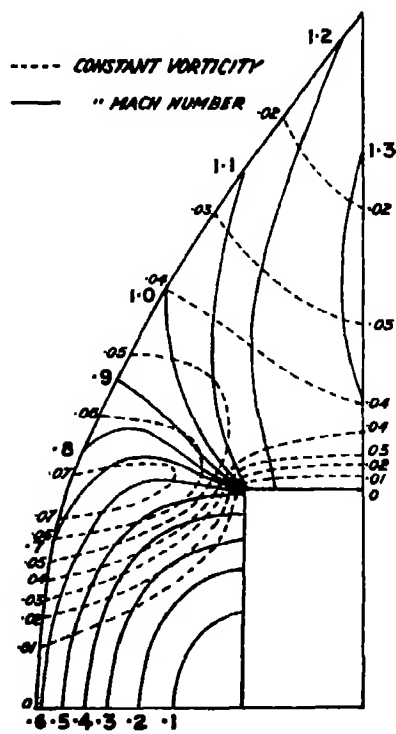
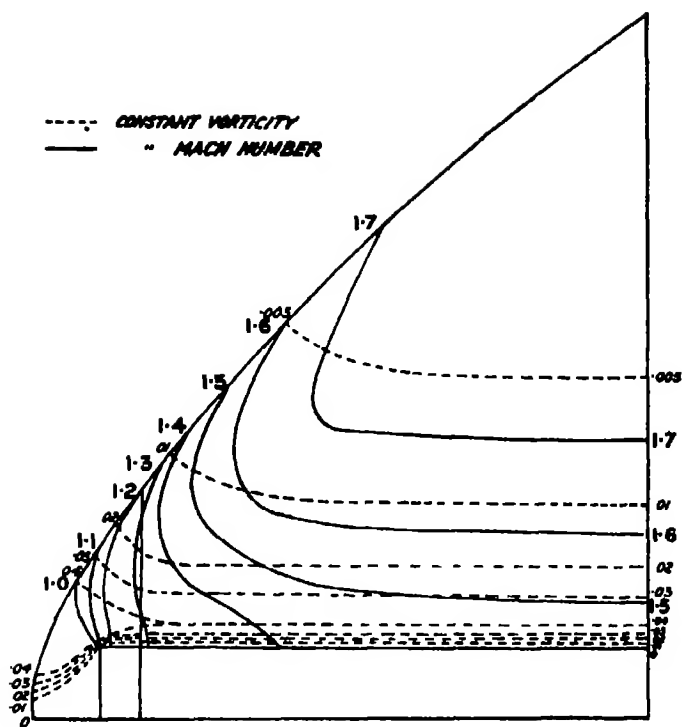


FIG. 3.

Using an initial network of mesh size $a = 1/40$, it is now possible to start applying equations (8) and (9) to a ψ -distribution in the region $ABCDEF$. The value of R is calculated from equation (8) for each node, and from a graph of R against χ the value of χ is found. The residuals F_0 are then calculated from equation (9). Again the difficulty of keeping R below its maximum value of 0.68 is encountered at nodes near the sonic line. The method employed to overcome this difficulty and in so doing to locate the sonic line is explained fully by Rutherford and one of the present authors (1951). In the reduction of residuals at nodes away from the sonic line, ψ is modified using the relaxation pattern. At nodes near the sonic line, however, a trial-and-error method of altering the stream function is again employed.

The part of the field containing the sonic line and the corner is examined in more detail using a finer net of mesh size $a = 1/80$. The complete field of flow showing the lines of constant Mach number is illustrated in figs. 4 and 5.



5. THE ROTATIONAL FIELD

The vorticity in the rotational field of flow behind the bow shock wave is given by Vazsonyi (1945),

$$\omega = \frac{\rho r}{R} \frac{\partial S}{\partial \psi}.$$

In non-dimensional form this equation becomes

$$\omega = \frac{\rho r}{\gamma v_r} \frac{1}{R} \frac{\partial S}{\partial \psi}, \quad (12)$$

where, with a slight change of notation, ω , ρ , r , and ψ now stand for the non-dimensional quantities $\left(\frac{h}{c_s}\right)$, $\left(\frac{\rho}{(\rho_s)_r}\right)$, $\frac{r}{h}$, and $\frac{\psi}{G}$ respectively. The quantity $\left(\frac{1}{\gamma v_r} \frac{1}{R} \frac{\partial S}{\partial \psi}\right)$ has a constant value along each streamline behind the shock. The non-dimensional vorticity is calculated from equation (12) at all nodes downstream of the shock, and lines of constant vorticity are shown in figs. 4 and 5.

The value of the vorticity on the shock wave is of course independent of the field downstream of the shock, and since $\frac{\rho r}{v_r}$ increases with increasing ψ along the shock, the vorticity is found to be a maximum after the point of inflexion on the graph of $(S_r - S_0)$ against ψ .

6. CONCLUSION

This paper shows that the mixed subsonic-supersonic rotational field behind an axially symmetric bow shock wave can be evaluated by means of an extension of Southwell's relaxation technique. It again demonstrates that a solution to a problem involving a mixed elliptic-hyperbolic type differential equation can be obtained using finite difference approximations. Because of the non-linear nature of the fundamental equations, it has not been possible to examine the finite difference approximations for stability and convergence (O'Brien, Hyman, and Kaplan, 1951).

As in the corresponding 2-dimensional problem, the Mach number well downstream of the shock wave varies from about 1.42 on the obstacle side to 1.80 on the first streamline not appreciably deflected by the shock, illustrating the importance of taking into account the rotational nature of the flow behind an axially symmetric bow shock wave. Due to the

increased curvature of the shock wave near the axis, the point on the shock at which the vorticity is a maximum is much nearer the axis than in the 2-dimensional case. Again the streamline starting there is a locus of high vorticity.

REFERENCES TO LITERATURE

- BUSEMANN, A., 1929. "Drücke auf Kegelförmige Spitzen bei Bewegung mit Überschallgeschwindigkeit", *Zeits. Angew. Math. Mech.*, ix, 496.
- DROUGGE, G., 1948. "The Flow around Conical Tips in the Upper Transonic Range", *The Aeronautical Research Institute of Sweden Report* 25.
- GREEN, J. R., and SOUTHWELL, R. V., 1943. "High Speed Flow of Compressible Fluid through a Two-Dimensional Nozzle", *Phil. Trans.*, A, CCXXXIX, 367-386.
- HANTZSCHE, W., and WENDT, H., 1942. "Supersonic Flow past Cones", *M.O.S. Translation*, 124.
- HOLDER, D. W., NORTH, R. J., and CHINNECK, A., 1949. "Observations of the Bow-Waves of Blunt-nosed Bodies of Revolution in Supersonic Airstreams", *A.R.C. Report* 12495.
- KOPAL, Z., 1947. "Tables of Supersonic Flow round Cones", *Tech. Rep.* 1, *Mass. Inst. Tech.*
- MACCOLL, J. W., and CODD, J., 1945. "Theoretical Investigations of the Flow around Various Bodies in the Sonic Region of Velocities", *M.O.S. Theoretical Research Report* 17.
- MITCHELL, A. R., 1951. "Application of Relaxation to the Rotational Field of Flow behind a Shock Wave", *Quart. Journ. Mech. Appl. Math.*, iv, 371-383.
- MITCHELL, A. R., and RUTHERFORD, D. E., 1951. "Application of Relaxation Methods to Compressible Flow past a Double Wedge", *Proc. Roy. Soc. Edin.*, A, LXIII, 139-154.
- O'BRIEN, G., HYMAN, M., and KAPLAN, S., 1951. "A Study of the Numerical Solution of Partial Differential Equations", *Journ. Math. Phys.*, xxix, 223-252.
- VAZSONYI, A., 1945. "On Rotational Gas Flows", *Quart. Appl. Math.*, iii, 29-37.

XXVIII.—A Molecular Sum Rule.* By **D. ter Haar**, Department of Natural Philosophy, University of St Andrews

(MS. received June 19, 1952. Revised MS. received August 27, 1952.
Read November 10, 1952)

SYNOPSIS

A derivation is given of a sum rule which may be useful in the discussion of line intensities of molecular spectra.

IN connection with a discussion of dissociation equilibria in interstellar space, Kramers and ter Haar (1946) came across the following sum rule. If λ is the molecular orbital quantum number, $\lambda=0, 1, 2, \dots$ corresponding respectively to $\sigma, \pi, \delta, \dots$ states (compare Mulliken, 1932), we have the selection rules $\Delta\lambda=0, \pm 1$, and if the oscillator strengths are denoted by f , we have the following relations:

$$\sum_{\Delta\lambda=-1} f = \frac{1-\lambda}{3}, \quad \sum_{\Delta\lambda=0} f = \frac{1}{3}, \quad \sum_{\Delta\lambda=+1} f = \frac{1+\lambda}{3}, \quad (1)$$

of which the special case of $\lambda=1$ was used by Kramers and ter Haar (1946) in their discussion of the CH spectrum.

It was originally planned to include the proof of (1) with other material, but this plan had to be abandoned. As the general relations (1) do not seem to be generally known while they may be of interest to molecular spectroscopists, and as the derivation of these relations is an excellent example of the elegant methods which were so characteristic of Kramers, I felt that it might be worth while to reproduce the derivation of (1) as it was outlined to me by Kramers when we were preparing our paper on the conditions in interstellar space.

The sum rule (1) can be applied whenever we are dealing with molecules where (a) the assignment of λ -values is justified, (b) we are interested in transitions involving one electron only. Examples are H_2 (one of the 1σ electrons), He_2^+ (the $2s\sigma$ electron), CH (the $2p\pi$ electron) and so on (see Mulliken, 1932, for the meaning of the symbols).

It may be remarked here that a similar derivation can be given of the sum rules applying to atomic quantum numbers and that (1) can be generalised to other cases.

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

Consider a system with cylindrical symmetry. The Hamiltonian operator \mathbf{H} will be of the form

$$\mathbf{H} = \Omega(\rho, s) - C \frac{1}{\rho^2} \frac{\partial^2}{\partial \chi^2} \quad (a)$$

where ρ , s and χ are cylindrical co-ordinates, and where

$$C = \hbar^2 / 8\pi^2 m \quad (3)$$

(\hbar : Planck's constant; m : electron mass), and where

$$\Omega(\rho, s) = -C \left[\frac{\partial^2}{\partial s^2} + \frac{\partial^2}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial}{\partial \rho} \right] + U(\rho, s), \quad (4)$$

with U the potential energy.

The stationary states of the molecule are characterised by two quantum numbers, n and λ , and the corresponding (normalised) characteristic functions are of the form

$$\phi_{\lambda, n}(\rho, s, \chi) = \frac{1}{\sqrt{2\pi}} \psi_{\lambda, n}(\rho, s) e^{i\lambda\chi}. \quad (5)$$

The energy levels E_n satisfy the equation

$$\mathbf{H}\phi_{\lambda, n} = E_n \phi_{\lambda, n}. \quad (6)$$

The completeness relation for the $\psi_{\lambda, n}$ has the form (Kramers, 1938)

$$\sum_n \psi_{\lambda, n}^*(\rho, s) \psi_{\lambda, n}(\rho', s') = [\rho\rho']^{-1} \delta(\rho - \rho') \delta(s - s'), \quad (7)$$

where $\delta(x)$ is Dirac's δ -function (Dirac, 1935).

The oscillator strength $f_{a; \beta}$ corresponding to a transition from a state a to a state β is given by the formula (Bethe, 1933)

$$f_{a; \beta} = (3C\epsilon^2)^{-1} [E_a - E_\beta] |P_{a\beta}|^2, \quad (8)$$

where ϵ is the electronic charge and where the $P_{a\beta}$ are the matrix elements of the electric polarisation.

In our case we are interested in the following three sums:

$$S_0 = \sum_n f_{\lambda, n; \lambda, n}, \quad S_{\pm} = \sum_n f_{\lambda, n; \lambda \pm 1, n}. \quad (9)$$

Expressing the operator of the electric polarisation in cylindrical co-ordinates and remembering the form of the characteristic functions (5), we see that we may use the following expressions for the polarisation operator:

$$\mathbf{P}_{\lambda, \lambda} = e\mathbf{s}, \quad \mathbf{P}_{\lambda, \lambda \pm 1} = \frac{1}{2}\epsilon(x \mp iy) = \frac{1}{2}\epsilon e^{\mp i\chi}. \quad (10)$$

From equations (8) to (10) we get

$$S_0 = -(3C)^{-1} \sum_{\pi} (E_{\pi} - E_n) \left| \int \phi_{\lambda, n}^* s \phi_{\lambda, \pi} dV \right|^2, \quad (11)$$

$$S_{\pm} = -(6C)^{-1} \sum_{\pi} (E_{\pi} - E_n) \left| \int \phi_{\lambda, n}^* (x \mp iy) \phi_{\lambda \pm 1, \pi} dV \right|^2, \quad (12)$$

where the volume element dV is given by the equation

$$dV = \rho d\rho ds d\chi = d\omega d\chi. \quad (13)$$

Using equation (6), introducing the $\psi_{\lambda, n}$ from equation (5), using equations (2) and (7) and integrating by parts, we get for S_0 ,*

$$\begin{aligned} S_0 &= -(3C)^{-1} \sum_{\pi} \iint dV dV' \phi_{\lambda, n}^* \phi'_{\lambda, n} s s' \phi_{\lambda, \pi}^* (H - E_n) \phi_{\lambda, \pi} \\ &= -(3C)^{-1} \iint d\omega d\omega' \psi_{\lambda, n}^* \psi'_{\lambda, n} s s' \left[\sum_{\pi} \psi_{\lambda, \pi}^* (\Omega + C\lambda^2/\rho^2 - E_n) \psi_{\lambda, \pi} \right] \\ &= -(3C)^{-1} \iint d\omega d\omega' \psi_{\lambda, n}^* \psi'_{\lambda, n} s s' [\Omega + C\lambda^2/\rho^2 - E_n] (\rho\rho')^{-1} \delta(\rho - \rho') \delta(s - s') \\ &= -(3C)^{-1} \int \rho d\rho ds s \psi_{\lambda, n} [\Omega + C\lambda^2/\rho^2 - E_n] s \psi_{\lambda, n}^*. \end{aligned} \quad (14)$$

Similarly we have for S_{\pm} the expression

$$S_{\pm} = -(6C)^{-1} \int \rho d\rho ds \rho \psi_{\lambda, n} [\Omega + C(\lambda \pm 1)^2/\rho^2 - E_n] \rho \psi_{\lambda, n}^*. \quad (15)$$

From equations (2) and (5) we have

$$[\Omega + C\lambda^2/\rho^2 - E_n] \psi_{\lambda, n} = 0, \quad (16)$$

and we get from equations (14) to (16) and (4) after some simplifications,

$$S_0 = -\frac{1}{2} \int \frac{\partial \psi_{\lambda, n}^*}{\partial s} s \psi_{\lambda, n} \rho d\rho ds, \quad (17)$$

$$S_{\pm} = -\frac{1}{2} \int \left[\mp 2\lambda \psi_{\lambda, n}^* \psi_{\lambda, n} + \frac{\partial \psi_{\lambda, n}^*}{\partial \rho} \rho \psi_{\lambda, n} \right] \rho d\rho ds. \quad (18)$$

From these last two equations and the fact that the $\psi_{\lambda, n}$ are normalised, equations (1) follow after integration by parts.

* Primes on a function denote that the argument of the function should be primed, $f' \equiv f(\rho', s', \chi')$.

REFERENCES TO LITERATURE

- BETHE, H., 1933. *Handb. Phys.*, xxiv, 431. Berlin (Springer).
DIRAC, P. A. M., 1935. *Quantum Mechanics*, 74. Oxford University Press.
KRAMERS, H. A., 1938. *Hand.- u. Jahrb. Chem. Phys.*, 1, 129. Leipzig, Akademische Verlagsgesellschaft.
KRAMERS, H. A., and TER HAAR, D., 1946. *Bull. astr. Inst. Netherlands*, x, 137 (especially § 5).
MULLIKEN, R. S., 1932. *Rev. Mod. Phys.*, iv, 1.

(Issued separately December 4, 1952)

XXIX.—The First Chemical Society, the First Chemical Journal, and the Chemical Revolution (Part II).* By James Kendall, M.A., D.Sc., LL.D., F.R.S., P.R.S.E.

(MS. received September 13, 1952. Read November 10, 1952)

SINCE the delivery of my presidential address (1) in July I have assembled an amount of supplementary information regarding "the Chemical Society instituted in the beginning of the Year 1785". This, together with a brief description of some other chemical societies of the revolutionary period, forms the basis of the present paper.

First of all, it will be expedient to furnish a complete list of the dissertations read before the Society during 1785–86 and included in the first volume of its *Proceedings*, appending short comments with respect to the communicators or their topics when anything of special interest arises.

DISSERTATIONS READ BEFORE THE CHEMICAL SOCIETY INSTITUTED
IN THE BEGINNING OF THE YEAR 1785

1. On Chemical Attraction, by Mr William Scott. It is very appropriate that the author of this first paper—he also contributed Paper II, on Fermentation—should be able to proclaim himself, in his inaugural dissertation presented to the Faculty of Medicine in June 1786, "Extraordinary Member and President for the Year of the Chemical Society of Edinburgh". Despite his name, Mr Scott was an Irishman. His thesis, entitled *De Acido atmospherico, sive aereo*, is in all probability one of the two—Mr Carmichaell's being the other—sent by Black to Lavoisier in 1790 as proof of the fact that Edinburgh students in general embraced his system. For it includes the following sentence:

Sed tametsi phlogiston repudiatur; omnia quæ sunt per regna naturæ dispersa, esse quandam mutationum chemicarum causam una voce conclamant; omnia, esse principium quod tantos inducere potest effectus, testantur; tale existere principium doctissimus Lavoisier nuper indicavit.†

This, coupled with the fact that Mr Scott makes it quite clear, in his paper presented "in the beginning of the year 1785", that he was even

* This paper was assisted in publication by a grant from the Carnegie Trust for the Universities of Scotland.

† But although phlogiston is rejected, all things which are scattered through the realms of nature testify unanimously that there is a certain cause of chemical changes; all things bear witness that there is a principle which is able to produce so many results; the most learned Lavoisier has recently demonstrated that such a principle does exist.

then sceptical about the existence of phlogiston (1, p. 357), advances still further the date at which it may be presumed that Joseph Black was endeavouring in his courses to make his pupils understand the new principles and explanations of the science of chemistry which Lavoisier had so happily invented (1, p. 354). Black must, indeed, have been their earliest advocate in Great Britain.

2. On Mercury, by Mr William Haliday (spelt Halliday in Ramsay's list). Mr Haliday is another Irishman who announces on the title-page of his graduation thesis *De Electricitate Medica* in 1786 that he is an Extraordinary Member and President for the Year of the Chemical Society. Evidently this Society, like the Royal Medical Society founded in 1737, adopted the custom of electing several presidents (usually four) annually (2). Mr Haliday's paper on Mercury contains a description of the different results obtained by subliming *Mercurius precipitatus per se* (our modern mercuric oxide) in vessels exposed to the action of the air, when "it forms crystals of a beautiful deep red colour like rubies", and by distilling it in vessels perfectly closed, when "it becomes flowing mercury again". He sagely concludes:

Thus is cleared up a seeming contradiction to be met with among the chemists, some of them asserting that the Calx is not reducible without the addition of Phlogiston, others of no less character denying this. The precipitate per se is a true Calx, as it is combined with pure Air and has acquired weight by the operation.

Some later remarks are not so happily worded, but in a subsequent communication (Paper 16) Mr Haliday also discloses that he inclines towards the new theory of combustion.

3. On Nitre, by Mr Samuel Black. Yet another Irishman who presented an inaugural dissertation, *De Ascensu Vaporum spontaneo*, to the Faculty of Medicine in 1786 and declares himself President for the Year of the Chemical Society on its title-page. It is to be hoped that Mr Samuel Black was no kinsman of his famous namesake, for he hesitates to pronounce any preference between the old and the new theories of combustion (1, p. 357), neatly evading the issue by stating, "I shall therefore leave the Society to adopt that one which may appear most agreeable to right reason and to known chemical laws". And he closes his communication with the philosophical remark:

When we see then how much these great and respectable men have imposed on themselves by indulging that propensity which mankind so universally discovered—inquiring into ultimate causes—let us learn wisdom from their errors. Let us not be carried away by theory any farther than we have data to support us.

Mr Black also contributed Paper 12, on Evaporation.

4. On Vitriolic Acid, by Mr Henry Johnston. He begins by saying that every person who is a member of this Society ought to contribute something for its support. He does not entertain any idea that his contribution contains anything new, and he is perfectly correct in that opinion, since he is an out-and-out phlogistonist. Sulphur, for him, is still a compound of vitriolic acid with phlogiston. Mr Henry Johnston is also the author of Paper 14.

5. On Opium, by Mr William Johnston. Also the author of Paper 20. His contributions, like those of his precursor Henry, are purely descriptive.

6. On Tartar, by Peter Gernon. Peter Gernon is the only original member of the Society whose name does not appear on the list of students registered in Joseph Black's class of Chemistry. It is possible that his name has been incorrectly transcribed on Ramsay's sheet, or incorrectly deciphered by Ramsay, since several errors in other names will be noted in the course of this enumeration, but I have not located any likely alternative.

7. Some Remarks on Arsenic, by Mr Robert Ross. No remarks; another phlogistonist.

8. Chemical Account of Alkaline Substances, by Mr L. L. Van Meurs (L. van Meurs in Ramsay's list). This "amteemann batavis" apologises for his want of ability to express himself in a foreign language, but he has written by far the longest (47 pages!) and perhaps one of the dullest papers in the whole volume.

9. On Lead, by Mr Biker M'Donald (spelt Bicker McDonald in Ramsay's list). A fourth Hibernian to style himself Extraordinary Member and President for the Year of the Chemical Society of Edinburgh in his thesis presented to the Faculty of Medicine in 1786. Also the communicator of Paper 13.

10. On Latent Heat, by Joshua Parr. This paper naturally teems with references to Joseph Black's pioneer research on the subject, carried out when he was Lecturer in Chemistry at Glasgow University twenty-five years previously, with the co-operation of his two most famous pupils, James Watt and William Irvine. So great was Black's aversion to publicity that this work was never printed during his lifetime, although he read an account of it to a literary and philosophical society that met at the University on April 23, 1762. Only four years after his death did Robison include a full description of it, prepared from Black's lecture notes and from a student's manuscript thereof, in his book *Lectures on the Elements of Chemistry, by the late Joseph Black* (3). Joshua Parr also confesses that he has drawn his material directly from Black's lectures.

The general results obtained by Black were, however, well known

to his contemporaries through his correspondence and through Watt's inventions. It was by the application of Black's new ideas on ebullition, and with Black's encouragement and generous help, that Watt developed his fundamental idea of economizing power by a separate condenser in the steam-engine. "It was this", wrote Robison, "that spread the knowledge of the doctrine of latent heat, and the name of Dr Black." And McKie and Heathcote (4) have noted that it is established beyond all doubt that Lavoisier was already familiar with Black's theory and method in 1772, yet his memoirs of 1777 and 1780 covering the same field do not even mention Black's name. We cannot wonder that Robison was indignant; but Black remained imperturbable, as I have described in detail in my previous paper. The progress of chemistry, not personal glory, was his primary concern.

Mr Parr is also the author of Paper 17, on Crystallization.

11. On Fermentation, by Mr William Scott. Mr Scott flatly rejects the phlogistic explanation of fermentation:

I consider the existence of phlogiston as merely ideal, and I perfectly coincide with the illustrious Buffon when he says that this phlogiston of the chemists is a being of their method rather than of Nature. . . . The phlogistic doctrine is in direct contradiction to all the laws of gravity and matter.

He is much more sympathetic to Dr Lubbock's theory of the *Principium Sorbili*—"one universal principle diffused through all nature, which is the chief agent in all its great operations, as combustion, calcination, fermentation, etc."—but confesses:

This principle is liable to the same objections that I have made to phlogiston. It may be said that the *Principium Sorbili* is equally hypothetical as the other, and may equally be considered as existing only in imagination.

And although his final discussion remains rather confused, since he is "unwilling to hazard conjectures unaided by fact, and unsupported by experiments", he shows rare acumen in the remark:

From the experiments of Mr Cavendish and the reasoning of Mr Lavoisier, water appears to be a real compound.

12. On Evaporation, by Mr Samuel Black. After criticizing a number of earlier and "altogether inadequate" theories regarding evaporation, the author of this paper proceeds thus:

We come now to a much more pleasing and agreeable part of our task, to discuss the opinions of men whose names do honour to Science. . . . Among these our very learned Professor Dr Black holds first place.

There follows a long and interesting discussion of the work of Joseph Black on latent heat.

13. On Iron, by Mr Biker M'Donald. Under the title of this communication Mr M'Donald has copied the couplet:

How many dangers do Inviron
The man who meddles with cold Iron!
HUDIBRAS.

He contrives to avoid these dangers very adroitly, however, by keeping his paper almost entirely descriptive.

14. Some few remarks on Silver, submitted to the candid examination of the Chemical Society by their fellow Member, Henry Johnston.

15. On Water, by John Crumbie. This paper bears the postscript: Edin^r, 26th, Nov^r—1785; the only direct reference to Edinburgh in the whole volume.

Strangely enough, Mr Crumbie does not even mention Mr Cavendish's recent discovery of the decomposition of water into its component parts, and still entitles it an *element*. If Mr William Scott (see Papers I and II) occupied the President's chair at this meeting, I feel sure that he started the discussion with some very searching questions.

16. Some remarks on the Calcination of Metals, submitted to the Discussion of the Chemical Society by their obliged Servt., W. Haliday. The artistic title-page of this paper and a brief extract hinting that Mr Haliday subscribes to the new theory of combustion have already been included in my presidential address (I). He details the objections to the old hypothesis and discusses the "beautiful experiments of Mr Lavoisier" at great length, but reveals that he has not completely freed himself from the phlogiston yoke.

17. On Crystallization, by Mr Joshua Parr. The only extract worthy of quotation from this paper is the following:

The doctrine of minute atoms I shall not intrude on the Society, as I do not think it worth their notice.

18. On Nitrous Air, by Dr Anthony Mann. Dr Mann, another Irishman, graduated in the Faculty of Medicine in 1785. He quotes, but doubts at certain points, Priestley's phlogistic explanation of the component parts of nitrous air. He himself suggests that it contains water.

19. On Copper, by Mr John Gay.

20. On Sugar, by Mr William Johnston.

21. On Quick Lime, by Mr John Sedgwick.

22. On Gold, by Mr William Lecky. Mr Lecky's name does not

appear on Ramsay's list of the fifty-nine original members; he presumably joined the Society after its inception. Some new blood was obviously urgently needed, in view of the frequent repetition of contributors manifest in sessions of the Society immediately preceding. Mr Lecky evidently justified his admission, since his medical graduation thesis in 1787 indicates that he was then President for the Year, the fifth Irishman to gain that distinction.

23. On Phosphorus, by Mr James McIlwaine (spelt McElwaine in Ramsay's list). Yet another Hibernian, who also obtained his medical degree in 1787. After first stating Stahl's theory of the combustion of phosphorus—"still held in repute by most modern chemists"—and then describing Lavoisier's recent experiments, which show that the large increase in weight on combustion corresponds exactly with that of the air absorbed, Mr McIlwaine continues:

I would hope that the Society will forgive my not giving a decided opinion in favour of either of them, as I am not yet so well acquainted with either as to know which of them I would give my preference to, but this I will leave to be decided by my ingenious fellow-members.

What a pity it is that the discussion after the paper is not recorded!

24. On Zinc, by Mr Thomas Burnside. And an Irishman again, who graduated in the Faculty of Medicine in 1786.

25. On Lead, by Mr Thomas Gill. Our first Englishman, a medical graduate of 1787.

26. On Coal, by Mr Richard (spelt Archd. in Ramsay's list) Webb.

27. On Magnesia, by Mr S. Latham Mitchill. Mr Samuel Latham Mitchill, like Mr Lecky, is not in Ramsay's list, and must have been a new recruit to the Society. From the title-page of his inaugural dissertation for the degree of M.D. in 1786, we learn that he was an American. His later career may be summarized from an account by Edgar F. Smith (5):

In 1792, he was elected to the Chair of Chemistry and Natural History in Columbia College. He opposed, in a very friendly way, the views of Priestley on phlogiston, and was the first teacher of chemistry in America to use the nomenclature of Lavoisier [so Black's influence spread also into the New World]. He founded the *Medical Repository*, the first paper in America devoted to general, as well as medical, science. He was probably the first American to write on chemical philosophy. He became in course of years an active member of nearly all the learned societies of the world [including the Royal Society of Edinburgh]. He was a sort of human dictionary whose opinion was sought by all originators and inventors of every grade. His ingenious theory of the doctrines of septon and septic acid gave impulse to Sir Humphry Davy's vast discoveries [rather an exaggeration, since Davy's disproof (6) of Mitchill's "theory of contagion" was a very juvenile effort]. He was a polished orator, a versifier and a poet, a man of infinite humour and excellent fancy.

Two brief extracts from Mr Mitchill's paper may be cited:

(a) I am very glad to have on this subject an author of so much originality and accuracy as Dr Black to copy; his experiments upon Magnesia have thrown a new light upon this part of chemistry and are so decisive and numerous that they hardly leave anything for the experimenter to devise or the writer to invent.

(b) It has been thought by some practitioners most advisable to prescribe the Calcined Magnesia, because they thus prevent the flatulence and belching so apt to ensue when the fixed air is set loose in the stomach.

After which, it is only fitting to quote again from Edgar F. Smith:

His eccentricities furnished material for the wits of the day to fashion many a joke at his expense, over which no one laughed more heartily than himself.

28. On Sulphur, by Mr Alexander Stevens. A phlogistonist who, at the end of his paper, requests the aid of his fellow-members in its discussion: "as it is a task I find myself entirely inadequate to".

29. Camphor and Volatile Alkali, by Mr William Symonds. Mr Symonds makes some sarcastic remarks against those who have not yet seen sufficient reason to abandon the hypothesis of phlogiston.

30. An attempt to point out some of the Consequences which flow from Mr Cavendish's Discovery of the Component Parts of Water, by Mr Thomas Beddoes. Mr Thomas Beddoes is the most famous of all the original fellows of the Chemical Society. He was born in Shropshire in 1760, attended classes at Edinburgh but graduated as M.D. at Oxford in 1786, and directed the Pneumatic Institution at Clifton from 1798 until his death in 1808 (7). A rare opportunity is presented him in this paper, and he throws it away. Perhaps he should not be judged too harshly, for Cavendish himself drew false conclusions from his own discovery, and remained a believer in phlogiston until his death.

In general, Mr Beddoes is a straddler on the grand controversy of the period:

I think it is common to both parties to assert more than they can prove. . . . If two different hypotheses can be adjusted to any set of appearances, both become uncertain, and that in much higher ratio than two to one. Perhaps we may conclude without much danger of error that they are both false, or at least imperfect.

He expresses the highest admiration for Mr Cavendish's work:

Priestley, Lavoisier and Laudriani have borne their testimony to the accuracy of one, whose own authority stands little in need of confirmation, since of him almost alone among modern philosophers it may be truly said that all his experiments have stood the test of repetition.

Notwithstanding these direct experiments, many have felt and still feel a strong repugnance to admit the consequences of them, some perhaps because they have always been taught to look upon water as an uncompound body.

To these I can only say, that if they are sure that God has bestowed upon them faculties by which they can discover the properties of natural bodies without the aid of experience, they do right in persisting.

He believes that Mr Cavendish has given phlogiston a reprieve (he will not say whether short or lasting):

Many operations can now be connected with the existence of phlogiston, which had admitted of no explanation before. Indeed it seems to me that but for this discovery, phlogiston must have been totally abandoned. When phosphorus is burnt in vital air, and the whole of it disappears, what account could have been given of this, different from Mr Lavoisier's, if we did not know that the production of water will account both for the diminution of the air and the increase in weight of the consumed body.

And, before wandering off to more important problems, such as how a hen introduces air into an egg and a conjecture concerning the use of manure, he simplifies the difficulty he has discussed above of two conflicting hypotheses by volunteering a third, hinted at (so he says) by Dr Black, which involves the assumption that Mr Cavendish's analytical experiments are, after all, illusory. He adds naively:

If it shall appear that his [Mr Cavendish's] observations have been inaccurate or imperfect, I shall follow the light of truth, but I shall look back upon the delusion with an eye of regret.

It is not surprising that Dr Beddoes, after Humphry Davy had brought his Pneumatic Institution into world-wide repute by his experiments on nitrous oxide, frittered away the rest of his life advocating drastic reforms in education, diet, dress, children's toys, and innumerable other topics. He corresponded frequently with Joseph Black after he left Edinburgh (8). In April 1789 he informed Black that Dr Priestley had totally overthrown the French chemists; two years later he wrote, "I am glad to see your renunciation of the old Chemical Theory in the *Annales de Chymie*"; in 1798 he encouraged Davy to publish a puerile attack on Lavoisier's doctrines (6, pp. 10-13). "As little fitted for a Mentor as a weather-cock for a compass"—Davy never wrote a more apposite epigram.

31. Some account of the Theories of Combustion, of Heat, of Light, and of Colour, by Mr John Carmichaell (spelt Carmichel in Ramsay's list) It is significant that Mr Carmichaell is the first known Scot to contribute to the *Proceedings* of the Society; how noteworthy his contribution was has been discussed in detail in my previous paper (1, pp. 350-352).

32. On Dephlogisticated Air, by Mr George Kirkaldie. Mr Kirkaldie is a second Scot, who obtained his medical degree in 1786, but he is not of the calibre of Mr Carmichaell. It is far from being his intention to enter upon the phlogiston controversy, and he merely proposes some observations on its nature for the consideration of the Society. He

criticizes the views of Mr Cavendish with regard to dephlogisticated air at great length, but his own suggestions are even more unsatisfactory.

This paper concludes the first volume of the *Proceedings* of the Chemical Society instituted in the beginning of the Year 1785.

THE IRISH QUESTION

The predominance of Irishmen among the communicators is truly astonishing. Of the fourteen contributors whose nationality is known, eight are Irish (and five of these became presidents of the Society), two Scots (one also was elected president, and compensates in quality for the lack of quantity), two English, one Dutch, and one American. Several of the original members who failed to find their way into the pages of the *Proceedings* subsequently submitted inaugural dissertations in the Faculty of Medicine, thereby disclosing their country of origin. They are listed in alphabetical order below, with their year of graduation.

1. John Barrow, England, 1785.
2. James Donovan (spelt Donovan in Ramsay's list), Ireland, 1784.
3. Nicholas Elcock, Ireland, 1786.
4. Edward Fairtlough (spelt Fairclough in Ramsay's list), Ireland, 1785.
5. James Forster, Ireland, 1785.
6. Thomas Galley, England, 1785.
7. Samuel Macay (spelt Macoy in Ramsay's list), Ireland, 1785.
8. Cornelius Pyne, Ireland, 1785.
9. William Robertson, Scotland, 1786.
10. John Unthank, Ireland, 1784.
11. John Watson-Sproule (spelt Sprole in Ramsay's list), Ireland, 1787.

Examination of all the medical theses submitted at the University of Edinburgh during this period has revealed a number of additional graduates who announce their membership of the Chemical Society on the title-page of their dissertations. They also are listed in alphabetical order below. The numbering in this list starts with 62, since the original 59 members, plus Mr Lecky and Mr Latham Mitchill, have precedence.

62. William Allanby, England, 1788 (President for the Year).
63. Samuel Crumpe, Ireland, 1788 (President for the Year).
64. William Jones Evans, Ireland, 1788.
65. Alexander Jackson, Ireland, 1787.
66. Charles Johnston, Ireland, 1785.
67. William Saunders O'Halloran, Ireland, 1788.
68. Thomas Renwick, England, 1787.
69. Joseph Sherlock, Ireland, 1788.
70. James Short, Scotland, 1788.
71. John Usher, Ireland, 1785.
72. William Whitelaw, Ireland, 1786.

The final state of the poll, with the nationality of exactly one-half of the total membership established, is therefore: Ireland 24 (6 presidents); England 6 (1 president); Scotland 4 (1 president); Holland 1; America 1. Our idea of the typical Edinburgh student of the eighteenth century, trudging to and from his home during the Meal Monday holiday to replenish his sack of oatmeal, obviously needs a little recasting.

With the year 1788 the Society appears to come to a sudden close, for I have continued my examination of the title-pages of the Edinburgh inaugural dissertations as far as the end of 1794 without encountering a single further graduate who proclaims himself a member of the Edinburgh Chemical Society. Since every one of the original members took special pride in this distinction, none qualified to do so omitting to state the fact, it is unfortunately clear that the first Chemical Society in the world faded out in its fourth year.

The year 1789, however, is not entirely lacking in interest, for one medical graduate, John Benjamin Jachmann, breaks new ground, styling himself an Honorary Fellow of the Chemical Society of Glasgow! As his name suggests, this solitary intruder from the west is also not a native Scot; he claims Prussian nationality. Dr Jachmann must have been a man of some distinction, for he was elected a President of the Royal Medical Society in the same year (2, p. 317), but I must leave it to my Glasgow colleagues to discover more about him. It will be relevant here, nevertheless, to give a brief account of this Glasgow society, as well as of some others of the period.

OTHER EIGHTEENTH-CENTURY CHEMICAL SOCIETIES

The history of the Chemical Society of Glasgow was unearthed by Dr J. A. V. Butler, formerly Lecturer in Chemistry at the University of Edinburgh, while working at Princeton University on leave of absence in 1941 (9). The first Professor of Chemistry at Princeton—then the College of New Jersey—was John Maclean, M.D., a Glasgow graduate. His son, John Maclean, became the tenth President of Princeton, and printed for private circulation a memoir of his father in 1876. In that memoir it is stated:

At the University he was, while yet a lad, a member of the Chemical Society, a club which appears to have met at the University, with the permission of the College authorities, if not under the oversight of the Professors. The members submitted, for the consideration of the Society, papers and essays upon various matters connected with the object of their association, and some of these papers seem to have foreshadowed the eminence which the authors of them attained in after life, as proficient in the art of Chemistry.

Other members of this Society were William Couper, Charles Macintosh (the inventor of the waterproof coat; see, however, p. 397), Mr Candlish, Dr Tilloch, Dr Crawford (a chemist of some repute) (10), Mr John Wilson, Major Finlay, Mr Cruikshank, Mr Archer and Mr Monroe. One of seven papers which Maclean himself read before the Society, dated March 29 [1786?], contains a reference to Dr William Irvine, Joseph Black's former collaborator (see p. 387), Lecturer in Chemistry at Glasgow University, who presumably sponsored its activities. Irvine died in the summer of 1787, and in October of that year Dr A. Eason of Manchester remarked in a letter to Charles Macintosh: "How does your chemical society get on? I am afraid, badly, since poor Irvine is gone. However, you must keep it up."

The precise date of the institution of the Glasgow Chemical Society is not known. It was most probably early in 1786, with late 1785 as a possibility, so that it runs as a very close second to Edinburgh. The exact date of its demise is also uncertain, but it was probably this same Society to which Mr John Benjamin Jachmann belonged in 1789. For in the biographical notice of Dr John Thomson, Professor of Medicine in the University of Edinburgh, his son Dr William Thomson states (11):

At the beginning of the winter session 1788-89, Mr John Thomson went to Glasgow to attend the medical classes. Besides prosecuting the study of anatomy with ardour, he attended the lectures of Dr Cleghorn, who was a lecturer on chemistry in the college, an office which had been successively held by Cullen, Black, and Irvine. He also joined a chemical society, which contained several members who afterwards attained great eminence as practical chemists. The doctrines of Lavoisier had just been made known, and gave much interest to the proceedings of young and ardent cultivators of chemical science, among whom it may be supposed that they found a readier reception than among those who, before adopting the new doctrines, had previously to unlearn the old.

The last sentence of this extract might have been written by Joseph Black himself (1, p. 354). It is not surprising to find that John Maclean, soon after he arrived in Princeton, engaged in a controversy with Priestley on the phlogiston theory (6, p. 232).

A second chemical society, however, was founded in Glasgow in 1798; its president in 1800 was William Ramsay, grandfather of Sir William Ramsay. The minute book of this society for the period October 8, 1800 to March 25, 1801, is now in the possession of the Royal Philosophical Society of Glasgow (12). It consists almost entirely of a record of experimental work by the society, apparently carried out at its meetings, which were held generally more frequently than once a month.

Simultaneously we find Dr John Thomson (11, p. 16 of Introduction) in Edinburgh converting a Natural History Society, probably at the

Royal College of Surgeons, into a Chemical Society, of which Lord Brougham and Lord Lauderdale were members.

There were thus, before the end of the eighteenth century, no fewer than four chemical societies instituted in Scotland—a truly admirable manifestation of the clan spirit. It is, furthermore, not unreasonable to regard the Chemical Society of Philadelphia as a daughter society of that at Edinburgh in 1785. The Medical School of the College of Philadelphia (now the University of Pennsylvania) was instituted in 1765 under strong Edinburgh auspices, and the coat of arms of the University of Edinburgh is still to be seen above the entrance to one of its original buildings. John Morgan, who first taught chemistry there, and Benjamin Rush, who succeeded him in 1769 as the first full-time Professor of Chemistry in America, were both students of Joseph Black in Edinburgh. With such an intimate connection between the two centres, the existence of a chemical society at Edinburgh University, sponsored by Black, would certainly be a matter of common knowledge in the College of Philadelphia, and when James Woodhouse (then a young man of 22) founded the Chemical Society of Philadelphia in 1792 he may quite plausibly be pictured as following, consciously and deliberately, in the footsteps of Joseph Black.

At least two eighteenth-century chemical societies also existed in England, but little definite is known about them. One must have flourished in Manchester well previous to 1787, since Dr Eason's letter to Charles Macintosh of that date, quoted on p. 395, contains the sentence: "We intend another volume of our transactions by next spring." Whether this good intention was fulfilled we know not, but what a wonderful discovery the *first* volume would be! The only reference to the second society is in the *Journal* of John Playfair, Robison's successor as Professor of Natural Philosophy in the University of Edinburgh (13). It was brought to my attention by an Aberdonian, the late Professor J. C. Philip.

Chemistry is the rage in London at present [1782]. I was introduced by Mr B. Vaughan (with whom I became acquainted in Edinburgh while he studied at the University there) to a chemical society, which meets once a fortnight at the Chapter Coffee-house. Here I met Mr Whithurst, Dr Keir, Dr Crawford, and several others.

Playfair goes on to record how he also saw Dr Priestley, "who has made so great a figure in the world". Priestley was then "particularly engaged in some experiments to prove that inflammable air is the same thing with phlogiston", but Playfair was too canny a Scot to be convinced.

If only this society were not so nebulous, it might well deprive the Chemical Society of Edinburgh of its priority. My own surmise regarding it is as follows. It will be noted that the London society to which the Edinburgh professor was introduced had a suspiciously Caledonian flavouring—Mr B. Vaughan was certainly one of Black's students, and two of the other three members named by Playfair are obvious Scots. I suspect that it possibly constituted an offshoot—the London branch—of a hypothetical older Edinburgh Chemical Society, for Mr Thomas Beddoes opens his communication to the *Proceedings* of the "society instituted in 1785" (the very title insinuates a precursor instituted in an earlier year) with the cryptic words:

Mr President: This Society, as well as others of older date and greater name, has often witnessed the doubts and difficulties that divide and perplex the chemists, concerning the nature and product of phlogistic processes.

To what societies is Mr Beddoes here alluding? The only definite suggestion I have to make is the Royal Medical Society, of which Mr Beddoes was President in its forty-ninth session, 1785–86. The minute-books of the Society for this period have been lost, but it is known (2, p. 62) that in 1785 a committee was appointed to supervise the fitting up of a laboratory for chemical experiments and, although the main interests of the Royal Medical Society were not strictly chemical, it must have held some discussions on the chemical revolution and its medical implications.

Obviously there is plenty still to be discovered by the diligent historian. One point that has just reached my notice is that the Chemical Society of the University of Edinburgh was resuscitated in 1815 by James Syme. Alexander Miles (14) states:

When he [Syme] became a member of Dr Hope's class in the University, he, along with Robert Christison and a dozen other students, founded a chemical society which met once a week.

And Sir Robert Christison wrote long afterwards (2, p. 126):

It very nearly made a chemist, instead of a surgeon, of Syme. Before it came to an end by the dispersion of its members, he had begun to work at the subject of the solvents of india-rubber, and his inquiries ended in his discovering its solubility in coal-tar naphtha, and the waterproofing of cloth by means of this solution. He published his discovery at a very early age. Nevertheless Macintosh, the manufacturing chemist, reaped all the honour as well as the profit.

This, however, is a digression from the eighteenth century, to which period I now return.

THE DISSERTATIONS SENT BY BLACK TO LAVOISIER IN
OCTOBER 1790

The number of inaugural dissertations "in which chemical subjects were chosen" at Edinburgh during the period 1785-90 is much fewer than I anticipated, and there is little doubt that the two sent by Black to Lavoisier as proof that the students of the University of Edinburgh "in general embrace your system and begin to make use of the new nomenclature" were:

1. *De Acido atmosferico, sive aereo*: William Scott, June 1786.
2. *De Fermentatione*: John Carmichaell, September 1787. Extracts from these dissertations have already been quoted on p. 385 and in my previous paper (1, p. 356).

The only other possibilities are:

1. *De Duobus Aeris speciebus aquam gignentibus*: George Kirkaldie, September 1786. This, however, contains nothing of interest to Lavoisier.
2. *De Igne*: Hugh Gillam, September 1786. Mr Gillam was not a member of the Chemical Society, and his thesis is officially placed in the physics section. It does include a phrase, "Quorum primus init prælia Lavoisier, chemiæ peritissimus ipse, et cæteri sequuntur", which would have been pleasant for Lavoisier to read, but I do not think Black would send to Lavoisier a thesis issued from another department.
3. *De Compositione Acidi Sulphurici*: Alexander Purcell Anderson, September 1790. The date probably eliminates this, since it is unlikely that printed copies would be available. Lavoisier is mentioned as "vir ingeniosus", but the theoretical aspects of the topic are scarcely touched.

The translation of Black's letter to Lavoisier in the *Annales de Chimie* (15) carries a footnote regarding the two inaugural dissertations sent therewith: "On en donnera incessamment l'extrait dans ce Journal." A careful search of succeeding issues, however, failed to find these promised summaries. It must be assumed that they disappeared in the turmoil of the Revolution.

The paucity of chemical theses submitted at Edinburgh in this period presumably springs from Black's increasing age and infirmity. One other Edinburgh adherent of Lavoisier's doctrines, however, deserves brief notice here, although I have not traced him as a student of Joseph Black. His name is Robert Kerr, and a letter from him to Lavoisier, written on January 21, 1791, is among those recently discovered by Dr Douglas McKie in the *Archives de France* (see 1, pp. 353-354).

In this letter (16) Mr Kerr first refers to the discussions which the Royal Society held in 1788 on the subject of phlogiston. There is no record in the *Transactions* of the Society as to what transpired at the five sessions devoted thereto, but Mr Kerr informs Lavoisier that "Sir James Hall Baronet had given our Edinburgh Royal Society a very ingenious detailed account of your doctrines which had shaken the phlogistic faith of many and even made several converts, amongst whom I was". Struck with the elegant simplicity of Lavoisier's theory, Mr Kerr immediately set about translating his *Elémens de Chymie* "in the hopes that it might prove an useful present to my countrymen". This altruism (or good business instinct) was properly rewarded, for although he hazarded a large edition, his bookseller now advises him that he must speedily prepare for reprinting. He accordingly presents Lavoisier with a copy of the translation, and requests instructions of any alterations or additions he would wish to have made in the next edition.

Lavoisier's answer has not been preserved, but it is interesting to note that four other editions followed between 1793 and 1802.

JOSEPH BLACK

This paper can most aptly close with a tribute to Joseph Black, the sponsor of the First Chemical Society, the inspirer of the contributors to the First Chemical Journal and, as now appears, in all probability the first chemist in Great Britain to join in the chemical revolution. How revered this grand old man of chemistry was by his colleagues is evinced by Lavoisier's letter to him (1, p. 353). How stimulating he was to his students may be judged by the following eulogy from the pen of Lord Brougham:

I have heard the greatest understandings of the age giving forth their efforts in its most eloquent tongue—have heard the commanding periods of Pitt's majestic oratory—the vehemence of Fox's declamations—have followed the close-compacted chain of Grant's pure reasoning—been carried away by the mingled fancy, epigram, and argumentation of Plunket; but I should without hesitation prefer, for mere intellectual gratification . . . to be once more allowed the privilege which I enjoyed a half century ago of being present while the first philosopher of his age was the historian of his own discoveries, and be an eye-witness of those experiments by which he had formerly made them, once more performed with his own hands.

Fellows will forgive me, I am sure, for mentioning the fact that I am a direct descendant, in the scientific sense, of Joseph Black, since the Chair of Chemistry at the University of Edinburgh has passed without interruption from professor to student for two centuries. I could not boast nobler ancestry.

REFERENCES TO LITERATURE

- (1) KENDALL, 1952. *Proc. Roy. Soc. Edin.*, A, LXIII, 346.
- (2) GRAY, 1952. *History of the Royal Medical Society*, Edinburgh University Press, 315.
- (3) ROBISON, 1803. *Lectures on the Elements of Chemistry, by the late Joseph Black*, Constable & Co., 1.
- (4) MCKIE and HEATHCOTE, 1935. *The Discovery of Specific and Latent Heats*, Edward Arnold & Co., 47.
- (5) SMITH, 1914. *Chemistry in America*, D. Appleton & Co., 148-150.
- (6) KENDALL, 1939. *Young Chemists and Great Discoveries*, G. Bell & Sons, 11.
- (7) STOCK, 1811. *Life of Thomas Beddoes, M.D.*, Murray.
- (8) RAMSAY, 1914. *The Life and Letters of Joseph Black, M.D.*, Constable & Co., 93-98.
- (9) BUTLER, 1942. *Journ. Chem. Educ.*, XIX, 43.
- (10) KENT, 1950. *An Eighteenth-Century Lectureship in Chemistry*, Jackson, Son & Co., 145.
- (11) THOMSON, 1859. *An Account of the Life, Lectures and Writings of William Cullen, M.D.*, William Blackwood & Sons, 8 (Introduction).
- (12) WILSON, 1937. *Ann. Sci.*, II, 451.
- (13) PLAYFAIR, 1822. *The Works of John Playfair*, Constable & Co. (Appendix).
- (14) MILES, 1918. *The Edinburgh School of Surgery before Lister*, Black, 175.
- (15) BLACK, 1791. *Annales de Chemie*, VIII, 225.
- (16) MCKIE, 1949. *Notes and Records of the Royal Society of London*, VII, 13.

(Issued separately December 4, 1952)

INDEX

- Abel's Series, Asymptotic Validity, by A. J. Macintyre and Sheila S. Macintyre, 222-231.
- Aitken (A. C.). Studies in Practical Mathematics: V. On the Iterative Solution of a System of Linear Equations, 52-60.
- Studies in Practical Mathematics.
- VI. On the Factorization of Polynomials by Iterative Methods, 174-191.
- Studies in Practical Mathematics.
- VII. On the Theory of Methods of Factorizing Polynomials by Iterated Division, 326-335.
- Artificial Holograms and Astigmatism, by G. L. Rogers, 313-325.
- Astigmatism, Artificial Holograms and, by G. L. Rogers, 313-325.
- Asymptotic Validity of Series—Abel's Series, by A. J. Macintyre and Sheila S. Macintyre, 222-231.
- β -disintegration Energies: Heavy β -emitters, by N. Feather, 242-256.
- Bhattacharyya (A.). Unbiased Statistics with Minimum Variance, 69-77.
- Bow Shock Wave, The Rotational Field behind a, in Axially Symmetric Flow, using Relaxation Methods, by A. R. Mitchell and Francis McCall, 371-380.
- Brownian Motion, Simple Model based on a Generalization of the Classical Random-walk Problem, by G. Klein, 268-279.
- Cambi (Enzo). The Simplest Form of Second-Order Linear Differential Equation, with Periodic Coefficient, having Finite Singularities, 27-51.
- Chemical Journal, The First, by J. Kendall, Part I, 346-358; Part II, 385-400.
- Revolution, by J. Kendall, Part I, 346-358; Part II, 385-400.
- Society, The First, by J. Kendall, Part I, 346-358; Part II, 385-400.
- CHENG (K. C.). See Green (H. S.) and Cheng (K. C.).
- Circular Cone, Normal Penetration of a Thin Elastic-Plastic Plate by a Right, by J. W. Craggs, 359-370.
- Clebsch-Aronhold Symbols and the Theory of Symmetric Functions, by H. W. Turnbull and A. H. Wallace, 155-173.
- Compressible Flow past a Double Wedge, Application of Relaxation Methods to, by A. R. Mitchell and D. E. Rutherford, 139-154.
- Continuant Determinants arising in Physics and Chemistry, by D. E. Rutherford, 232-241.
- Co-variance, On the Estimation of Variance and, by E. H. Lloyd, 280-289.
- Craggs (J. W.). The Normal Penetration of a Thin Elastic-Plastic Plate by a Right Circular Cone, 359-370.
- Daniels (H. E.). The Statistical Theory of Stiff Chains, 290-311.
- Determinants, Some Continuant, in Physics and Chemistry: II, by D. E. Rutherford, 232-241.
- Difference-differential Equations—the Linear, with Constant Coefficients, by E. M. Wright, 18-26.
- Differential Equation, Second-Order Linear, with Periodic Coefficient, having Finite Singularities, by Enzo Cambi, 27-51.
- Diffraction Microscopy, Experiments in, by G. L. Rogers, 193-221.
- Double Wedge, Application of Relaxation Methods to Compressible Flow past a, by A. R. Mitchell and D. E. Rutherford, 139-154.
- Elastic-Plastic Plate, Normal Penetration by a Right Circular Cone, by J. W. Craggs, 359-370.
- Electrodynamics, Reciprocity Theory of, by H. S. Green and K. C. Cheng, 105-138.
- Electron-capture Process, Sargent Diagram for, by N. Feather, 242-256.
- Equation, Solution of a Functional, by A. H. Read, 336-345.
- Factor Analysis, A Further Note on a Problem in, by D. N. Lawley, 93-94.
- Factorization of Polynomials by Iterative Methods, by A. C. Aitken, 174-191.
- Factorizing Polynomials by Iterated Division, by A. C. Aitken, 326-335.
- Feather (N.). The Sargent Diagram for the Electron-capture Process, and the Disintegration Energies of Heavy β -emitters, 242-256.
- First Chemical Society, the First Chemical Journal, and the Chemical Revolution, by J. Kendall, Part I, 346-358; Part II, 385-400.
- Fréchet (Maurice). Les Transformations asymptotiquement presque périodiques discontinues et le lemme ergodique. (Première Note), 61-68.
- Functional Equation, The Solution of a, by A. H. Read, 336-345.
- Gabor Diffraction Microscopy: Experiments in Diffraction Microscopy, by G. L. Rogers, 193-221.
- Green (H. S.) and Cheng (K. C.). The Reciprocity Theory of Electro-dynamics, 105-138.

- Holograms, Artificial Holograms and Astigmatism, by G. L. Rogers, 313-325.
 — Experiments in Diffraction Microscopy, by G. L. Rogers, 193-221.
 Houston (R. A.). A Measurement of the Velocity of Light, 95-104.
 Hypothesis, Adventures of an, by James Kendall, 1-17.
- Iterated Division, Factorization of Polynomials by, by A. C. Aitken, 326-335.
 Iteration, Solution of Linear Equations by, by A. C. Aitken, 52-60.
 Iterative Methods, Factorization of Polynomials by, by A. C. Aitken, 174-191.
- Kendall (James). The Adventures of an Hypothesis, 1-17.
 — The First Chemical Society, the First Chemical Journal, and the Chemical Revolution, Part I, 346-358; Part II, 385-400.
 Klein (G.). A Generalization of the Classical Random-walk Problem, and a Simple Model of Brownian Motion based thereon, 268-279.
- Lawley (D. N.). A Further Note on a Problem in Factor Analysis, 93-94.
 Light, A Measurement of the Velocity of, by R. A. Houston, 95-104.
 Linear Equations, Iterative Solution of, by A. C. Aitken, 52-60.
 Line Intensities; A Molecular Sum Rule, by D. ter Haar, 381-384.
 Lloyd (E. H.). On the Estimation of Variance and Co-variance, 280-289.
- McCall (Francis). See Mitchell (A. R.) and McCall (Francis).
 Macintyre (A. J.) and Macintyre, Sheila (S.). Theorems on the Convergence and Asymptotic Validity of Abel's Series, 222-231.
 Macintyre (Sheila S.). See Macintyre (A. J.) and Macintyre (Sheila S.).
 Microscopy, Experiments in Diffraction, by G. L. Rogers, 193-221.
 Mitchell (A. R.) and McCall (Francis). The Rotational Field behind a Bow Shock Wave in Axially Symmetric Flow, using Relaxation Methods, 371-380.
 Mitchell (A. R.) and Rutherford (D. E.). Application of Relaxation Methods to Compressible Flow past a Double Wedge, 139-154.
- Normal Penetration of a Thin Elastic-Plastic Plate by a Right Circular Cone, by J. W. Craggs, 359-370.
- Parallel Planes in a Riemannian V_n , by H. S. Ruse, 78-92.
 Penetration of a Thin Elastic-Plastic Plate by a Right Circular Cone, by J. W. Craggs, 359-370.
- Polynomials, Factorization of, by Iterative Methods, by A. C. Aitken, 174-191.
 — Factorizing by Iterated Division, by A. C. Aitken, 326-335.
 Prime Number Theorem, Elementary Proof of the, by E. M. Wright, 257-267.
- Random-walk Problem, A Generalization of the Classical, by G. Klein, 268-279.
 Read (A. H.). The Solution of a Functional Equation, 336-345.
 Reciprocity Theory of Electrodynamics, by H. S. Grech and K. C. Cheng, 105-138.
 Relaxation Methods, Application of, to Compressible Flow past a Double Wedge, by A. R. Mitchell and D. E. Rutherford, 139-154.
 — Rotational Field behind a Bow Shock Wave in Axially Symmetric Flow, by A. R. Mitchell and Francis McCall, 371-380.
 Riemannian V_n , Parallel Planes in a, by H. S. Ruse, 78-92.
 Rogers (G. L.). Artificial Holograms and Astigmatism, 313-325.
 — Experiments in Diffraction Microscopy, 193-221.
 Ruse (H. S.). Parallel Planes in a Riemannian V_n , 78-92.
 Rutherford (D. E.). Some Continuant Determinants arising in Physics and Chemistry: II, 232-241.
 — See Mitchell (A. R.) and Rutherford (D. E.).
- Spectroscopy: A Molecular Sum Rule, by D. ter Haar, 381-384.
 Statistical Theory of Stiff Chains, by H. E. Daniels, 290-311.
 Statistics, Unbiased, with Minimum Variance, by A. Bhattacharyya, 69-77.
 Stiff Chains, Statistical Theory of, by H. E. Daniels, 290-311.
 Sum Rule, A Molecular, by D. ter Haar, 381-384.
 Symmetric Functions, treated by Symbolic Methods, by H. W. Turnbull and A. H. Wallace, 155-173.
- ter Haar (D.). A Molecular Sum Rule, 381-384.
 Theorem, Prime Number, Elementary Proof of, by E. M. Wright, 257-267.
 Transformations (Les) asymptotiquement presque périodiques discontinues et le lemme ergodique. (Première Note), by Maurice Fréchet, 61-68.
 Turnbull (H. W.) and Wallace (A. H.). Clebsch-Aronhold Symbols and the Theory of Symmetric Functions, by H. W. Turnbull and A. H. Wallace, 155-173.
- Unbiased Statistics with Minimum Variance, by A. Bhattacharyya, 69-77.

- Variance and Co-variance, On the Estimation of, by E. H. Lloyd, 280-289.
- Velocity of Light, A Measurement of the, by R. A. Houstoun, 95-104.
- Wallace (A. H.). *See* Turnbull (H. W.) and Wallace (A. H.).
- Wright (E. M.). The Stability of Solutions of Non-linear Difference-differential Equations, 18-26.
- The Elementary Proof of the Prime Number Theorem, 257-267.
- Zone Plate: Experiments in Diffraction Microscopy, by G. L. Rogers, 193-221.

I.A.R.L. 75

**INDIAN AGRICULTURAL RESEARCH
INSTITUTE LIBRARY, NEW DELHI.**

[illegible]

GIPNLK-H-40 I.A.R.I.-29-4-55-15,000



**INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.**

I. A. R. I. 6.

MGIPC-81-51 AR/57-3-4-58-5,000.